



European  
Commission

# **OPZET EN OPDRACHTVERLENING VAN CONTRAFEITELIJKE EFFECTONDERZOEKEN**

**EEN PRAKTISCHE LEIDRAAD VOOR BEHEERAUTORITEITEN  
VAN HET ESF**

**EUROPESE COMMISSIE**

Directoraat-generaal Werkgelegenheid, Sociale Zaken en Inclusie  
Directoraat G — Fondsen: programmering en uitvoering  
Eenheid G5 — Betere regelgeving

*Contact:* Linda Adamaite

*E-mail:* [EMPL-G5-UNIT@ec.europa.eu](mailto:EMPL-G5-UNIT@ec.europa.eu) of [Linda.Adamaite@ec.europa.eu](mailto:Linda.Adamaite@ec.europa.eu)

*Europese Commissie*

*B-1049 Brussel*

# **OPZET EN OPDRACHTVERLENING VAN CONTRAFEITELIJKE EFFECTONDERZOEKEN**

EEN PRAKTISCHE LEIDRAAD VOOR  
BEHEERAUTORITEITEN VAN HET ESF

Manuscript voltooid in september 2021

De editie van 2021 van “Opzet en opdrachtverlening van contrafeitelijke effectonderzoeken — Een praktische leidraad voor beheerautoriteiten van het ESF” is geschreven door Jochen Kluve (Humboldt-Universität zu Berlin) en Andrea Naldini en Marco Pompili (Isméri Europa). Deze editie is een bewerking van de leidraad van 2013 (ISBN 978-92-79-28238-6; DOI 10.2767/94454) die werd opgesteld door Stephen Morris (Policy Evaluation and Research Unit, Manchester Metropolitan University), Herta Tödting-Schönhofer (Metis GmbH, Wenen) en Michael Wiseman (George Washington Institute of Public Policy).

Dit document is opgesteld in opdracht van de Europese Commissie, maar geeft slechts de standpunten van de auteurs weer. De Europese Commissie is niet aansprakelijk voor de gevolgen die voortvloeien uit hergebruik van deze publicatie. Meer informatie over de Europese Unie is beschikbaar op internet (<http://www.europa.eu>).

Luxemburg: Bureau voor publicaties van de Europese Unie, 2021

© Europese Unie, 2021



Het beleid inzake hergebruik van documenten van de Europese Commissie is geregeld bij Besluit 2011/833/EU van de Commissie van 12 december 2011 betreffende het hergebruik van documenten van de Commissie (PB L 330 van 14.12.2011, blz. 39).

Tenzij anderszins vermeld, is hergebruik van dit document toegestaan krachtens de Creative Commons Attribution 4.0 International (CC BY 4.0)-licentie (<https://creativecommons.org/licenses/by/4.0/>). Dit betekent dat hergebruik is toegestaan, mits de bron correct wordt aangegeven en eventuele wijzigingen worden vermeld.

Voor het gebruik of de reproductie van onderdelen die geen eigendom zijn van de Europese Unie, kan het nodig zijn de rechthebbenden rechtstreeks om toestemming te vragen.

Voorblad: © Shutterstock, 2021

PDF ISBN 978-92-76-40725-6

doi:10. 10.2767/02762

KE-02-21-992-EN-N

## Inhoud

<b>INLEIDING: ACHTERGROND EN DOEL VAN DE LEIDRAAD .....</b>	<b>5</b>
<b>HOOFDSTUK 1. CONCEPT EN BENADERING.....</b>	<b>11</b>
1.1. DE ESSENTIE VAN CONTRAFEITELIJKHEID.....	11
1.2. WAT IS HET BELANG VAN CONTRAFEITELIJKE EFFECTONDERZOEKEN? .....	12
1.3. WAAROM ZIJN CONTRAFEITELIJKE EFFECTONDERZOEKEN IN TECHNISCH OPZICHT EEN UITDAGING? .....	14
1.4. OVERZICHT VAN CIE-MODELLEN EN -METHODEN.....	15
1.5. HOE KAN EEN CIE IN EEN BREDER EVALUATIEKADER WORDEN GEÏNTEGREERD? .....	16
<b>HOOFDSTUK 2. PRAKTISCHE OVERWEGINGEN BIJ HET OPZETTEN VAN EEN CIE.....</b>	<b>24</b>
2.1. DE INTERVENTIES VOOR EEN EFFECTONDERZOEK SELECTEREN .....	27
2.1.1. <i>Interventies voor effectonderzoeken prioriteren</i> .....	29
2.1.2. <i>Interventies selecteren die geschikt zijn voor een contrafeitelijke aanpak</i> .....	31
2.2. DE EVALUATIEVRAGEN EN UITKOMSTVARIABLEN.....	36
2.2.1. <i>Wat zijn de doelstellingen van de interventie?</i> .....	36
2.2.2. <i>Waarom dient de evaluatie?</i> .....	36
2.3. GEGEVENS WAARMEE DE CONTROLEGROEP KAN WORDEN BEPAALD EN DE UITKOMSTVARIABLEN KUNNEN WORDEN GEMETEN.....	41
2.3.1. <i>Zijn de juiste gegevens beschikbaar of kunnen die beschikbaar komen?</i> .....	41
2.3.2. <i>Hoe moet de “behandelingsgroep” worden bepaald?</i> .....	54
2.3.3. <i>Aandachtspunten bij het vaststellen van een controlegroep</i> .....	55
2.3.4. <i>Welke soorten gegevenskwesties moeten in het evaluatieschema aan de orde komen?</i> .....	59
2.3.5. <i>Wat zijn de belangrijkste beperkingen bij het analyseren van gegevens en resultaten?</i> .....	61
2.3.6. <i>Een checklist om de voorbereiding en haalbaarheid van het CIE te verifiëren</i> .....	64
2.4. TOE TE PASSEN CIE-METHODE .....	65
2.5. TIJDSHEMA EN BEGROTING .....	66
2.5.1. <i>Welke middelen zijn beschikbaar?</i> .....	66
2.5.2. <i>Wanneer moet de interventie worden geëvalueerd?</i> .....	71
2.6. UITVOERING VAN HET CIE .....	74
2.6.1. <i>Selectie van de beoordelaar</i> .....	74
2.6.2. <i>Toezicht op CIE's</i> .....	76
2.6.3. <i>Rapportage</i> .....	76
2.6.4. <i>Benutting van de resultaten</i> .....	78
<b>HOOFDSTUK 3. HOE KAN DE JUISTE METHODE VOOR DE UITVOERING VAN EEN CIE WORDEN GEKOZEN? .....</b>	<b>81</b>
3.1. RANDOMISATIE — DE EXPERIMENTELE BENADERING .....	81
3.2. NIET-GERANDOMISEERDE OF QUASI-EXPERIMENTELE OPZETTEN.....	85
3.2.1. <i>Doel- en controlegroep zonder randomisatie</i> .....	85
3.2.2. <i>Propensity score matching</i> .....	87
3.2.3. <i>Difference-in-differences</i> .....	89
3.2.4. <i>Regressie-discontinuïteitsanalyse</i> .....	92
3.2.5. <i>Instrumentele variabelen</i> .....	95
<b>HOOFDSTUK 4. HOE NU VERDER?.....</b>	<b>102</b>

<b>EEN AGENDA VOOR CIE'S</b> .....	<b>102</b>
4.1. VERBETERING VAN HET INZICHT BIJ BELANGHEBBENDEN.....	102
4.2. CAPACITEITSONTWIKKELING .....	103
4.3. JURIDISCHE BELEMMERINGEN AANPAKKEN.....	106
4.4. OP WEG NAAR MEER TOEKOMSTGERICHTE BENADERINGEN.....	106
4.5. HET TOEPASSINGSGEBIED VAN CIE'S UITBREIDEN.....	107
<b>GLOSSARIA</b> .....	<b>113</b>
AFKORTINGEN.....	113
DEFINITIES.....	114
<b>BIJLAGEN</b> .....	<b>122</b>
BIJLAGE 1. LITERATUUR .....	122
BIJLAGE 2. VOORGESTELDE OPZET VOOR EEN CIE-CURSUS .....	126
BIJLAGE 3 CONTRAFAEITELIJKE EFFECTONDERZOEKEN — IN DE LEIDRAAD GENOEMDE VOORBEELDEN .....	127

## LIJST VAN TEKSTKADERS

TEKSTKADER 1 VOORBEELD VAN EEN CIE VOOR HET VERGELIJKEN VAN KOSTEN EN BATEN .....	23
TEKSTKADER 2 INTEGRATIE VAN CIE'S IN EEN BREDER KADER .....	27
TEKSTKADER 3 VRAGEN BIJ HET SELECTEREN VAN INTERVENTIES VOOR EEN CIE .....	28
TEKSTKADER 4 SOORTEN ESF-INTERVENTIES EN -DOELGROEPEN DIE HET VAAKST WORDEN GEKOZEN VOOR CONTRAFEITELIJKE EFFECTONDERZOEKEN .....	31
TEKSTKADER 5 CONTROLEGROEPEN AFBAKENEN .....	35
TEKSTKADER 6 VOORBEELDEN VAN GEGEVENS DIE VOOR CIE'S WORDEN GEBRUIKT .....	43
TEKSTKADER 7 VOORBEELDEN VAN GEÏNTEGREERDE DATABANKEN VOOR CIE'S .....	44
TEKSTKADER 8 EU-REGELGEVINGSKADER VOOR DE VERWERKING VAN PERSOONSGEGEVENS .....	48
TEKSTKADER 9 BESCHERMING EN UITWISSELING VAN GEGEVENS .....	53
TEKSTKADER 10 BELEIDSVRAGEN IN VERBAND MET EEN OPLEIDINGSPROGRAMMA .....	58
TEKSTKADER 11 DE NETTO-EFFECTEN INTERPRETEREN .....	58
TEKSTKADER 12 ONZEKERHEID BIJ DE INTERPRETATIE VAN DE RESULTATEN .....	63
TEKSTKADER 13 DE POOLSE ERVARING MET EVALUATIECONFERENTIES .....	80
TEKSTKADER 14 VOORBEELD VAN EEN GERANDOMISEERD ONDERZOEK VAN EEN ESF-PROJECT VOOR JONGEREN .....	84
TEKSTKADER 15 VOORBEELD VAN EEN EVALUATIE WAARBIJ GEBRUIK WORDT GEMAAKT VAN EEN MATCHINGMETHODE .....	89
TEKSTKADER 16 VOORBEELD VAN EEN EVALUATIE AAN DE HAND VAN EEN DID-METHODE .....	91
TEKSTKADER 17 VOORBEELD VAN EEN EVALUATIE WAARBIJ EEN RDD WORDT GEBRUIKT .....	95
TEKSTKADER 18 VOORBEELD VAN EEN STUDIE MET GEBRUIK VAN EEN INSTRUMENTELE-VARIABELENMETHODE .....	98
TEKSTKADER 19 VOORBEELD VAN EEN PROJECT TER VERSTERKING VAN DE CIE-CULTUUR EN -CAPACITEIT .....	105
TEKSTKADER 20 VOORBEELDEN VAN EVALUATIES OP ONDERWIJSGBIED .....	108
TEKSTKADER 21 VOORBEELD VAN DE BEOORDELING VAN EFFECTEN OP SOFT OUTCOMES IN DUITSLAND .....	111

## LIJST VAN FIGUREN

FIGUUR 1 VERSCHILLENDE EVALUATIETAKEN EN -SOORTEN .....	18
FIGUUR 2 WEERGAVE VAN EEN LOGISCH MODEL OF INTERVENTIELOGICA .....	19
FIGUUR 3 HOOFDVOLGORDE VAN CIE-ACTIVITEITEN .....	25
FIGUUR 4 MINIMAAL DETECTEERBARE EFFECTGROOTTEN (MDES) BIJ VERSCHILLENDE STEEKPROEFGROOTTEN .....	62
FIGUUR 5 VEREENVOUDIGD TIJDSHEMA VOOR DE RESULTATEN VAN EEN OPLEIDINGSPROGRAMMA .....	73
FIGUUR 6 GERANDOMISEERD GECONTROLEERD ONDERZOEK MET TWEE GROEPEN .....	83
FIGUUR 7 GESTILEERDE QUASI-EXPERIMENTELE OPZET MET BEHANDELINGS- EN CONTROLEGROEP .....	86
FIGUUR 8 WEERGAVE VAN DE PROPENSITY SCORE-METHODE .....	88
FIGUUR 9 WEERGAVE VAN DE DID-METHODE .....	91
FIGUUR 10 WEERGAVE VAN DE REGRESSIE-DISCONTINUÏTEITSMETHODE .....	93
FIGUUR 11 WEERGAVE VAN EEN INSTRUMENTELE-VARIABELENMETHODE .....	96

## LIJST VAN TABELLEN

TABEL 1 AANBEVOLEN INHOUD VAN EEN EVALUATIESHEMA .....	26
TABEL 2 SOORTEN GEGEVENS EN GEGEVENSBRONNEN .....	42
TABEL 3 STRUCTUUR VAN DE BELANGRIJKSTE KOSTEN VAN EEN CIE .....	70
TABEL 4 BASISINFORMATIE DIE MOET WORDEN OPGENOMEN IN EEN BLAD WAARIN HET CIE WORDT GEPRESENTEERD .....	78
TABEL 5 VERGELIJKING VAN DE HOOFDKENMERKEN VAN DE BELANGRIJKSTE CIE-METHODEN .....	99
TABEL 6 KENMERKEN VAN DE CIE'S DIE IN DE LEIDRAAD ALS VOORBEELD WORDEN GENOEMD .....	127

# Inleiding: achtergrond en doel van de leidraad

De programmeringsperiode 2021-2027 gaat van start tijdens de ingrijpende COVID-19-pandemie en de daaropvolgende economische crisis. Om bij te dragen aan het herstel van de nationale economieën en aan de groei van de werkgelegenheid heeft de Europese Unie (EU) haar financiële en beleidsmatige inspanningen opgevoerd. Het Europees Sociaal Fonds Plus (ESF+) vervult in dat opzicht een belangrijke rol door werklozen — en in het bijzonder werkloze jongeren en vrouwen — grootschalig hulp te bieden, door interventies ter bestrijding van armoede onder kinderen te ondersteunen, beter onderwijs te bevorderen en de sociale inclusie van achterstandsgroepen in de hele EU te stimuleren. De uiteenlopende doelstellingen van het ESF+ en de noodzaak om snel resultaten te boeken op het gebied van werkgelegenheid en sociale inclusie, nopen tot een doeltreffende toewijzing van middelen. Empirisch onderbouwde beleidsbenaderingen worden steeds belangrijker en evaluaties zijn een fundamenteel instrument om richting te geven aan het overheidsbeleid.

*Leidraad voor  
beheerautoriteiten  
met updates op  
grond van de  
ervaringen in de  
periode 2014-2020*

De Europese Commissie heeft in de periode 2014-2020 de evaluatiecapaciteit van de lidstaten en de beheerautoriteiten ondersteund en het gebruik van contrafeitelijke effectonderzoeken (*Counterfactual Impact Evaluations*, CIE's) aangemoedigd. De positieve resultaten die in de CIE's zijn geconstateerd, vormen tastbaar bewijs van het feit dat het ESF effecten sorteert die anders achterwege zouden zijn gebleven. Een aanzienlijk aantal CIE's werd eerst in de evaluatieplannen opgenomen om vervolgens tijdens de uitvoering van de ESF-programma's hun beslag te krijgen. Bij het opstellen en uitvoeren van de CIE's deden zich in veel gevallen problemen voor of ging het om op zichzelf staande exercities die niet in een systematisch evaluatiekader waren ingebed.

Deze leidraad is bestemd voor beheerautoriteiten en andere instanties die verantwoordelijk zijn voor de uitvoering van interventies en programma's die uit het ESF+ worden gefinancierd en is bedoeld als hulpmiddel bij de planning, opzet en opdrachtverlening van CIE's. In de leidraad is rekening gehouden met eerdere ervaringen en worden praktische adviezen gegeven met betrekking tot enkele belangrijke vragen waarover bij het opzetten van een CIE moet worden nagedacht. Deze leidraad is een update van de eerdere leidraad voor de periode 2014-2020 en besteedt meer aandacht aan de problemen die zich bij de praktische uitvoering van CIE's hebben voorgedaan. Desondanks komen de methodologische aspecten wel ter sprake. Waar mogelijk worden deze vereenvoudigd en verwerkt in geactualiseerde voorbeelden die zijn ontleend aan de ESF-evaluaties in de periode 2014-2020.

CIE's geven antwoorden op de cruciale vragen waarmee empirisch onderbouwde beleidsbeslissingen kunnen worden genomen: wat zijn de causale effecten van interventies en “wat werkt er”? Met een CIE wordt geprobeerd na te gaan of de gewijzigde situatie van de deelnemers aan een interventie daadwerkelijk kan worden toegeschreven aan de uit het ESF gefinancierde interventie en wat er vervolgens met de interventie is bereikt.

*Een CIE draait om  
de vraag: “wat werkt  
er?”*



Een goed uitgevoerd CIE levert empirische gegevens op over het “netto-effect”<sup>1</sup> van een interventie. Beleidsmakers kunnen dan uitsluiten dat er alternatieve verklaringen zijn voor de gewijzigde situatie van de deelnemers of voor de resultaten die eventueel zijn geconstateerd. Schattingen in CIE’s van het bestaan en de omvang van het netto-effect bevatten onvermijdelijk een bepaalde mate van onzekerheid, afhankelijk van de methodologische nauwkeurigheid en de beschikbare informatie. Met de empirische gegevens van CIE’s kunnen beleidsmakers de effectiviteit van interventies onderzoeken, interventies onderling vergelijken en de relatieve prestaties ervan beoordelen. Daarnaast leveren CIE’s belangrijke input voor kosten-batenanalyses en kostenefficiëntieanalyses.

Deze leidraad verschijnt in een tijd waarin het ESF+ voor ongekende uitdagingen staat. Gezien de enorme toename van de EU-middelen voor investeringen en werkgelegenheid van het pakket NextGenerationEU, is het cruciaal dat beleidsmakers de effecten van de interventies waarvoor zij verantwoordelijk zijn meten en begrijpen. Om het herstel te bespoedigen en de sociale ongelijkheid terug te dringen, moeten overheidsmiddelen worden vrijgemaakt voor interventies die productiever en effectiever zijn. De verantwoordelijken voor de uitbetaling van de ESF+-middelen moeten hun keuzes onderbouwen door aan te tonen dat hun interventies effectief en van meerwaarde voor de burger zijn. Dit kan het beste door een groot aantal kwalitatief hoogwaardige CIE’s uit te voeren.

Het ESF+ is het belangrijkste Europese instrument voor de ondersteuning van werkgelegenheid en sociale inclusie. In de programmeringsperiode 2014-2020 werd uit het ESF bijna 125 miljard EUR besteed aan beleidsmaatregelen op het gebied van arbeidsmarktactivering, onderwijs en sociale inclusie. De uitvoering daarvan verliep via operationele programma’s in de 28 lidstaten. Zoals vastgelegd in Verordening (EU) nr. 1303/2013 houdende gemeenschappelijke bepalingen “worden evaluaties verricht om de kwaliteit van het ontwerp en de uitvoering van programma’s te verbeteren en de doeltreffendheid, de doelmatigheid en het effect ervan te beoordelen”.

In de programmeringsperiode 2021-2027 zal het onderzoek naar de prestaties en resultaten worden voortgezet<sup>2</sup>. Daarvoor is het nodig dat de bestaande systemen en capaciteit voor monitoring en evaluatie worden versterkt, waaronder de voorschriften inzake gegevensverzameling. De evaluatieplannen blijven verplicht en er komt meer nadruk te liggen op effectonderzoeken. Omdat er verschillende methoden bestaan om de effecten te inventariseren van de concrete acties die vanuit het ESF+ zijn ondersteund, moeten de beheerautoriteiten bepalen met welke methode of combinatie van methoden optimaal aan de regelgevingseisen kan worden voldaan. Bij een gedegen kwantificering van interventie-effecten hoort ook een contrafeitelijk effectonderzoek.

*Resultaatgerichtheid  
en hoogwaardige  
evaluaties*

Een belangrijk kenmerk van de nieuwe regelgeving is dat de nadruk ligt op krachtige prestaties en resultaatgerichtheid. Hoogwaardige

<sup>1</sup> Het netto-effect is het verschil tussen het totale effect (of bruto-effect) en de resultaten die zonder de interventie zouden zijn behaald. Als de effectiviteit van de interventie achterblijft bij de marktdynamiek kan het netto-effect ook negatief zijn. Zie blz. 9 voor meer informatie.

<sup>2</sup> VERORDENING (EU) 2021/1060 VAN HET EUROPEES PARLEMENT EN DE RAAD van 24 juni 2021 houdende gemeenschappelijke bepalingen inzake het Europees Fonds voor regionale ontwikkeling, het Europees Sociaal Fonds Plus, het Cohesiefonds, het Fonds voor een rechtvaardige transitie en het Europees Fonds voor maritieme zaken, visserij en aquacultuur en de financiële regels voor die fondsen en voor het Fonds voor asiel, migratie en integratie, het Fonds voor interne veiligheid en het Instrument voor financiële steun voor grensbeheer en visumbeleid.

evaluatiestrategieën en -technieken zijn cruciaal voor het achterhalen van essentiële informatie die elke lidstaat laat zien welke interventies wel en welke interventies niet “werken”. Het is essentieel dat de kwaliteit van de evaluaties omhooggaat en dat er betrouwbare empirische gegevens over de meerwaarde worden ontwikkeld.

Het uitgangspunt bij het vergaren van empirische gegevens over de effectiviteit van beleidsinterventies is in wezen eenvoudig. Dit vereist:

- dat wordt vastgesteld welk probleem moet worden aangepakt;
- dat wordt bepaald met welke instrumenten het probleem wordt aangepakt;

en een formule waarmee de instrumenten en de resultaten aan elkaar worden gekoppeld.

Om een financieringsregeling of -instrument te kunnen evalueren, is het van cruciaal belang dat er duidelijke en meetbare indicatoren bestaan voor zowel de gehanteerde inputs als de probleemgerelateerde outputs en resultaten. Het is gebruikelijk om streefdoelen te formuleren voor zowel de outputs als de resultaten en om de feitelijke resultaten tegen die streefdoelen af te zetten. Via monitoring worden de inputs, outputs en resultaten gevolgd en wordt feedback aan het management geleverd. In de onderliggende interventietheorie wordt vaak op tussentijdse resultaten gewezen. De monitoring kan ook op die resultaten worden gericht. Om vanuit dit uitgangspunt terecht te komen bij het antwoord op de vraag of een bepaalde interventie werkt, is echter nog niet zo eenvoudig.

*De noodzaak van resultaten...*

De bestaande evaluatiepraktijk heeft behoefte aan extra methoden waarmee veel hardere gegevens over de netto-effecten van interventies kunnen worden verkregen. Resultaatmeting is een kwestie van rekenschap afleggen over de gebruikte financiële middelen. In een CIE staat de fundamentele vraag naar de effectiviteit van een interventie centraal. Met een CIE wordt weliswaar geprobeerd om een oorzakelijk verband tussen interventies en resultaten te leggen, maar er kunnen aanvullende theoriegebaseerde methoden en procesevaluaties nodig zijn om de onderliggende oorzakelijke mechanismen in kaart te brengen en ervoor te zorgen dat de effecten die op een bepaalde locatie zijn gerealiseerd een empirische aanleiding bieden om datzelfde beleid ook elders toe te passen.

*...en empirische gegevens over de netto-effecten*

In de programmeringsperiode 2014-2020 hebben alle lidstaten en beheerautoriteiten een evaluatieplan vastgesteld waarin de evaluatiedoelstellingen, activiteiten, organisatorische aspecten en producten worden beschreven. De plannen voorzien in diverse soorten evaluaties, zoals doorlopende evaluaties van programma's, thematische evaluaties gericht op de beantwoording van specifieke evaluatievragen of evaluaties waarin bepaalde prioritaire assen of investeringsprioriteiten van een operationeel programma centraal staan. Uit een analyse<sup>3</sup> van de evaluatieplannen met betrekking tot 177 ESF-programma's is gebleken dat er voor deze programmeringsperiode ongeveer 132 contrafeitelijke effectonderzoeken op stapel staan (stand van zaken van eind 2018). Dat aantal getuigt van vooruitgang en van het feit dat er in de

*Ervaringen van de lidstaten met CIE's*

---

<sup>3</sup> Zie [Ismeri Europa — Ecorys — Institute for Employment Studies, 2019](#)

programmeringsperiode 2014-2020 meer aandacht voor contrafeitelijke methoden bestaat dan in de periode 2007-2013<sup>4</sup>.

Volgens het projectregister van de evaluatiehelpdesk<sup>5</sup> werden er tot en met juni 2021 1 795 evaluaties afgerond, waarvan er 1 001 betrekking hadden op ESF/YEI- of ESF/YEI- en EFRO-programma's (respectievelijk 675 en 326). Van de 1 001 evaluaties die het ESF betroffen, ging het in 323 gevallen om effectonderzoeken en werd bij 106 daarvan een contrafeitelijke methode gehanteerd: 82 van de 234 effectonderzoeken hadden betrekking op ESF/YEI-programma's en 24 van de 89 effectonderzoeken betroffen ESF/YEI- en EFRO-programma's<sup>6</sup>.

*Het gebruik van CIE's neemt toe...*

In de beginjaren van de huidige programmeringsperiode stond in de contrafeitelijke effectonderzoeken van ESF/YEI- en/of EFRO-programma's de vorige programmeringsperiode centraal, terwijl sinds 2018 in bijna alle CIE's de effecten van de interventies uit de periode 2014-2020 worden beoordeeld<sup>7</sup>. 20 van de 27 lidstaten hebben contrafeitelijke effectonderzoeken naar de programma's uitgevoerd, onder meer naar het ESF, hetgeen er nogmaals op wijst dat deze aanpak nu op grotere schaal wordt toegepast dan in het verleden.

Hoewel de methodologische aanpak van CIE's per lidstaat verschilt, werd de contrafeitelijke methode vooral gebruikt om de effecten van beleidsmaatregelen op het gebied van arbeidsmarktactivering (opleiding, prikkels, hulp bij het zoeken naar werk, werkervaring) te beoordelen; op andere gebieden gebeurde dat in veel mindere mate, zoals bij interventies waarmee onderwijsinstellingen en studenten werden ondersteund of bij sociaal beleid, zoals interventies ter bestrijding van armoede of sociale uitsluiting.

Qua methodologie wordt het vaakst gebruikgemaakt van *propensity score matching* en spelen andere methoden een minder prominente rol. In CIE's staan vaak de effecten op kort- of middellange termijn (6 of 12 maanden) centraal. Naar de langetermijneffecten wordt minder vaak onderzoek gedaan.

Uit de analyse van CIE's die in het kader van het evaluatiehelpdesk-project is uitgevoerd, komen onder meer de volgende beperkingen naar voren: de mensen in de controlegroepen lijken onvoldoende op de mensen in de ondersteunde groepen; de betrouwbaarheid van de gegevens is gering, bijvoorbeeld wanneer die gegevens afkomstig zijn uit enquêtes of interviews achteraf en wordt aangenomen dat de respondenten zich gebeurtenissen nauwkeurig kunnen herinneren; een kleine steekproefomvang die statistisch niet-significante resultaten oplevert enz. In sommige evaluatieverslagen ontbreekt het aan voldoende informatie over de methodologische stappen, keuzen en beperkingen, terwijl die informatie noodzakelijk is om niet alleen de opdrachtgevers maar ook anderen iets te kunnen vertellen over de effecten van de onderzochte maatregelen en om hun meer inzicht in die effecten te verschaffen.

*...maar er zijn nog volop beperkingen*

<sup>4</sup> Zie [Bratu C. et al., 2014](#)

<sup>5</sup> Met het door DG REGIO en DG EMPL gefinancierde helpdeskproject wordt informatie verzameld over de evaluaties die sinds 1 januari 2015 op de website van de beheerautoriteiten zijn gepubliceerd. Een netwerk van nationale deskundigen brengt de evaluaties in kaart; de evaluaties worden samengevat en beoordeeld op de kwaliteit en de betrouwbaarheid van de bevindingen. [https://ec.europa.eu/regional\\_policy/en/policy/evaluations/member-states/](https://ec.europa.eu/regional_policy/en/policy/evaluations/member-states/)

<sup>6</sup> Van de 106 werden er 90 door het helpdeskproject onderzocht en samengevat.

<sup>7</sup> 29 van de 33 CIE's die tot 2018 werden uitgevoerd hadden betrekking op de voorgaande programmeringsperiode.

Kortom, de uitvoering van CIE's vergt technische deskundigheid en politieke wil, ondanks de vooruitgang van de afgelopen jaren. In deze leidraad wordt een lans voor CIE's gebroken en worden enkele punten benoemd die door de beheerautoriteiten moeten worden opgepakt om de uitvoering van die CIE's tot een succes te maken. Om de kwaliteit van effectonderzoeken naar een hoger plan te kunnen tillen, gaat de aandacht niet alleen uit naar de praktische aspecten van CIE's maar ook naar de bredere vraagstukken die eventueel om een antwoord vragen. Deze leidraad bevat een aantal elementaire aanbevelingen waarmee de beheerautoriteiten hopelijk hun voordeel kunnen doen. Een van de belangrijkste boodschappen is echter dat de uitvoering van CIE's een zorgvuldige planning van de gegevensvoorbereiding vergt (met name van administratieve gegevens) alsook duidelijke doelstellingen enz., zodat eventuele problemen in de uitvoeringsfase of kwalitatief laagwaardige evaluaties kunnen worden voorkomen.

*Een leidraad voor vakmensen*

Als houvast voor de lezer bevat de leidraad tevens voorbeelden van evaluaties en praktijken. Hoewel de in deze leidraad gepresenteerde voorbeelden en praktijken hoofdzakelijk afkomstig zijn uit de evaluaties die door de evaluatiehelpdesk zijn verzameld, zijn ook andere bronnen geraadpleegd, met name de evaluaties die door het Gemeenschappelijk Centrum voor onderzoek (Centre for Research on Impact Evaluation (CRIE)) zijn uitgevoerd en, in mindere mate, wetenschappelijke publicaties. De voorbeelden en ervaringen in de leidraad hebben zo veel mogelijk betrekking op interventies in het kader van het Europees Sociaal Fonds.

De leidraad bestaat uit vier hoofdstukken.

In **hoofdstuk 1** wordt ingegaan op de aard en het belang van CIE's. In dit hoofdstuk staat een inleidend overzicht van de CIE-methoden en wordt het onderscheid tussen experimentele en quasi-experimentele methoden onderstreept. Verder worden daarin de technische en praktische problemen benoemd die bij de uitvoering van een CIE overwonnen moeten worden. Er wordt op hoofdlijnen stilgestaan bij het soort beleidsvragen waarop CIE's een antwoord kunnen bieden alsook bij het verband tussen CIE-methoden en andere evaluatiemethoden (zoals theoriegebaseerde benaderingen, procesevaluaties en efficiëntieanalyses).

*Een opbouw die de evaluatiestappen laat zien*

In **hoofdstuk 2** wordt ingegaan op een reeks vragen waarover de beheerautoriteiten zich bij het opzetten en uitvoeren van CIE's moeten buigen. De leidraad schetst een aantal van de belangrijkste problemen waarmee de ontwikkelaars van CIE's vaak worden geconfronteerd en doet een paar aanbevelingen om die problemen te verhelpen. De vragen fungeren als houvast voor mensen die opdracht willen geven tot een contrafeitelijk effectonderzoek naar een uit het ESF gefinancierde interventie.

CIE's leveren hoogwaardige empirische gegevens op over de effectiviteit van fondsen, maar alleen als er sprake is van goede planning en adequate uitvoering. Daarom moeten de beheerautoriteiten bij het verlenen van een evaluatieopdracht op een aantal belangrijke punten letten. Hoe de beheerautoriteiten deze punten precies afhandelen en in welke volgorde, wordt bepaald door praktische aspecten en door de institutionele regelingen ter plaatse in de lidstaat. In de leidraad worden enkele van deze belangrijke punten benadrukt en onder de aandacht van de beheerautoriteiten gebracht. De volgorde van de belangrijkste punten correspondeert met de zes stappen

van het evaluatieproces: 1) de te beoordelen concrete actie(s) kiezen; 2) de evaluatievragen en uitkomstvariabelen vaststellen; 3) gegevens analyseren om de controlegroep te bepalen en de uitkomstvariabelen te meten; 4) de CIE-methode kiezen; 5) het tijdschema en de begroting vaststellen; 6) de evaluatie uitvoeren. Aan de beschikbaarheid en bescherming van gegevens wordt speciale aandacht besteed, want dit aspect kan ernstig afbreuk doen aan de mogelijkheden voor het uitvoeren van CIE's.

In **hoofdstuk 3** wordt gekeken naar stap 4 (de CIE-methode kiezen) en staan de kenmerken van de belangrijkste methoden voor contrafeitelijke analyses centraal. Vijf methoden worden in het bijzonder besproken: de experimentele methode of randomisatie, propensity score matching, verschil-in-verschillen, regressie-discontinuïteitsanalyse en instrumentele variabelen. Dat deel van de leidraad is niet bedoeld om de lezer te trainen in het gebruik van CIE-methoden, maar om hem of haar bewust te maken van de sterke en zwakke punten van elke methode en om te laten zien wanneer de ene methode geschikter is dan de andere.

**Hoofdstuk 4** gaat in op meer algemene vraagstukken rond de ontwikkeling van CIE in de toekomst. Daarbij gaat het onder meer om de capaciteitsontwikkeling, zowel bij beheerautoriteiten intern (beleidsmakers en ambtenaren) als in kringen van onderzoek en wetenschap en bij adviesbureaus in de lidstaten, die nodig is om CIE's met succes te kunnen uitvoeren. Dat deel gaat ook in op de noodzaak om het hoofd te bieden aan juridische belemmeringen omtrent de toegang tot gegevens, om de opzet van de CIE's van ESF+-programma's te moderniseren, zodat ook het beleid op het gebied van sociale inclusie en onderwijs wordt bestreken, en op de noodzaak van effectramingen met betrekking tot *soft outcomes* (niet rechtstreeks meetbare uitkomsten) en de toepassing van vormen van meta-evaluatie.

Samengevat gebeurt in deze leidraad het volgende: 1) er wordt een lans gebroken voor CIE's; 2) de belangrijke stappen voor een succesvolle uitvoering van CIE's worden benoemd; en 3) CIE's worden tot een onmisbaar onderdeel van het ESF+-landschap bestempeld. Het uiteindelijke doel is dat het ESF+ een grotere bijdrage levert aan het welzijn van de burgers in Europa.

De auteurs zijn erkentelijk voor de waardevolle steun en hulp die zij bij het opstellen van deze leidraad hebben gekregen van de leden van Eenheid G5 — Betere regelgeving van DG Werkgelegenheid, en in het bijzonder van Linda Adamaite Jeannette Monier en Maria José Cueto Faus. Verder bedanken de auteurs Andrea Pisano en Ernesto Belisario voor hun bijdrage aan het gedeelte over gegevensbescherming. Ook Paweł Hess wordt bedankt voor zijn steun en bijdrage.

# Hoofdstuk 1. Concept en benaderingen

In dit hoofdstuk komen fundamentele vragen aan de orde over de aard en het doel van de contrafeitelijke methode bij het beoordelen van de causale effecten van interventies. Meer concreet bevat het hoofdstuk een uiteenzetting over de essentie van een contrafeitelijk effectonderzoek, vooral in het kader van de soorten concrete acties die worden medegefinancierd uit het ESF+. Ook wordt het verband tussen contrafeitelijke en andere evaluatiemethoden onderzocht en wordt het belang — en met name het hedendaagse belang — van CIE's besproken. Onderzocht wordt op welke beleidsvraagstukken CIE's een antwoord kunnen bieden. Voordat zij in hoofdstuk 3 uitvoerig aan de orde komen, wordt een beknopt overzicht gegeven van de belangrijkste contrafeitelijke methoden die relevant zijn bij het evalueren van interventies die uit het ESF+ zijn medegefinancierd.

## 1.1. De essentie van contrafeitelijkheid

Met een CIE wordt geprobeerd vast te stellen wat de netto-effecten van interventies zijn. Het onderscheidende kenmerk van CIE's is dat zij bedoeld zijn als onderbouwing van de hypothese dat een bepaalde interventie een concreet resultaat oplevert. Dat wil zeggen dat dit concrete resultaat louter aan de interventie kan worden toegeschreven. Dat kan met een CIE omdat daarmee de interventie wordt afgezonderd en alternatieve verklaringen voor het waargenomen resultaat worden uitgesloten.

Dat met een CIE alternatieve verklaringen kunnen worden uitgesloten, is te danken aan het feit dat een CIE is gebaseerd op het idee van "contrafeitelijkheid". Dat is het antwoord op de vraag: "Hoe zouden de uitkomsten in kwestie er hebben uitgezien als de interventie niet was uitgevoerd?" Voor een goed begrip van het concept "contrafeitelijkheid" kan het nuttig zijn om te kijken naar een voorbeeld dat omwille van de duidelijkheid eenvoudig is gehouden: een werkloze persoon doet mee aan een opleidingsprogramma dat bedoeld is om de werkgelegenheid te bevorderen. Om het effect van de opleiding op die persoon te kunnen bepalen, worden bij de contrafeitelijke methode twee mogelijke resultaten geformuleerd<sup>8</sup>. Het eerste resultaat is de arbeidsstatus van de cursist nadat hij of zij de opleiding heeft doorlopen. Dit is het waargenomen resultaat voor de cursist. Het tweede mogelijke resultaat is de arbeidsstatus van de cursist indien hij of zij de opleiding niet zou hebben gevolgd, aannemende dat de overige omstandigheden onveranderd zijn gebleven. In dat geval wordt dat tweede resultaat het "contrafeitelijke resultaat" genoemd. Het effect van de opleiding op de individuele cursist bestaat uit het verschil tussen het waargenomen en het contrafeitelijke resultaat. Dit is het oorzakelijke netto-effect van de opleiding op de betrokkene. De omstandigheden of voorwaarden die tot het waargenomen en het contrafeitelijke resultaat hebben geleid, verschilden slechts op één

*Contrafeitelijkheid*

<sup>8</sup> In [Holland P., 1986](#) wordt het causale model van de "hypothetische uitkomsten" uitvoeriger besproken.

punt: de betrokkene heeft een opleiding gevolgd. Het eventuele verschil tussen die twee resultaten moet dus wel het gevolg zijn van de opleiding op de arbeidsstatus van de betrokkene.

In werkelijkheid worden bij personen die een interventie ondergaan geen contrafeitelijke resultaten waargenomen, en dat kan ook niet. Een CIE is echter primair bedoeld om overtuigende schattingen te maken van de contrafeitelijke resultaten voor groepen personen of ondernemingen die interventies ondergaan die uit het ESF worden medegefinancierd. Het effect wordt daarom bijvoorbeeld uitgedrukt als het verschil tussen de gemiddelden of verhoudingen van enerzijds de gemiddelde waargenomen waarden en anderzijds de “geschatte” contrafeitelijke waarden. Een CIE wordt veelal ingezet om de resultaten van een interventie (een maatregel of een instrument) voor de entiteiten (personen, kmo’s enz.) die van die interventie gebruik hebben gemaakt te vergelijken met de resultaten voor een groep die de interventie niet heeft ondergaan. In de CIE-terminologie wordt onderscheid gemaakt tussen de “behandelde groep” of de “behandelingsgroep” en de “controlegroep”, die in alle opzichten zo veel mogelijk moet lijken op de behandelingsgroep (met uitzondering van de verleende behandelingen). De controlegroep fungeert als basis voor de schattingen van de contrafeitelijke resultaten, met speciale aandacht voor waargenomen en niet-waargenomen verschillen in de kenmerken van die twee groepen. Het is ook mogelijk om een aantal verschillende behandelingen met elkaar te vergelijken. Daartoe worden geschikte eenheden onderworpen aan een reeks behandelvarianten (zoals andere uit het ESF gefinancierde behandelingen of interventies die uit andere bronnen zijn gefinancierd), wordt een aantal behandelingsgroepen samengesteld en worden de resultaten onderling vergeleken en/of vergeleken met de resultaten voor een niet-behandelde controlegroep.

*Afbakening van behandelingsgroepen...*

Wanneer de controlegroep geen behandeling ondergaat, luidt de evaluatievraag die moet worden gesteld: “Wat is het effect van de interventie ten opzichte van de situatie waarin geen hulp of steun wordt verleend?” Omgekeerd geldt dat wanneer de resultaten van de behandeling in kwestie worden vergeleken met de resultaten van een andere behandeling, de te stellen evaluatievraag luidt: “Wat is het effect van de betreffende interventie ten opzichte van de situatie waarin mensen een nauwkeurig omschreven alternatieve interventie ondergaan?” In veel gevallen kan een CIE zodanig worden vormgegeven dat het CIE op een van deze fundamentele vragen een antwoord geeft. De keuze voor de te beantwoorden vraag wordt bepaald door de prioriteiten van beleidsmakers en door de praktische beperkingen van het evaluatiemodel.

*...en controlegroepen*

Wanneer twee verschillende behandelingen met elkaar worden vergeleken, moet er sprake zijn van een duidelijk omschreven en in beleidsmatig opzicht zinnig onderscheid tussen die behandelingen.

## 1.2. Wat is het belang van contrafeitelijke effectonderzoeken?

CIE’s leveren belangrijke informatie op over de netto-effecten van interventies, geven een schatting van de omvang van de effecten, geven de richting daarvan aan (negatief of positief) en bieden statistische maatstaven voor de onzekerheid. Met behulp van CIE’s kan het

veronderstelde causale verband tussen de interventie en de resultaten waarop het ontwerp van de interventie is gebaseerd, worden gevalideerd of verworpen. Deze effectmetingen kunnen voor een aantal doeleinden worden gebruikt: om aan te tonen dat het geld van de belastingbetaler transparant wordt besteed en dat daar rekenschap over wordt afgelegd; om beleidsbeslissingen te nemen die op duidelijk bewijs berusten; om lessen te trekken uit de resultaten voor diverse regio's binnen hetzelfde land of voor verschillende lidstaten.

Diegenen die verantwoordelijk zijn voor de interventies en ervoor moeten zorgen dat hun programma's financiering blijven aantrekken, hebben veel belang bij het bevorderen van CIE's. Zo kunnen zij laten zien dat hun programma een goede prijs-kwaliteitverhouding heeft en zowel voor de deelnemers als voor de samenleving als geheel meetbare voordelen oplevert. Voor de empirische gegevens van CIE's zal met name belangstelling leven onder de verantwoordelijken voor de toewijzing van middelen. De beheerautoriteiten willen graag laten zien dat hun programma's inderdaad "werken". Om dit op een overtuigende manier te kunnen doen, zullen zij opdracht tot kwalitatief hoogwaardige CIE's moeten geven.

*Transparantie en verantwoording*

Dankzij deze eigenschappen voorzien CIE's de beleidsmakers die belast zijn met de toewijzing van middelen aan verschillende interventies van belangrijke informatie. Besluiten over de financiering van eventuele interventies worden genomen in situaties waarin sprake is van beperkte middelen. In zo'n situatie hebben beleidsmakers behoefte aan solide gegevens over de effecten en de kosteneffectiviteit van programma's, zodat zij de beschikbare middelen met optimaal resultaat kunnen inzetten. Na een onderzoek naar de kosteneffectiviteit van een interventie of de uitvoering van een volledige kosten-batenanalyse kunnen de in het CIE geraamde netto-effecten in de beoordeling van de relatieve efficiëntie van interventies uitvoeriger worden gepresenteerd.

*Onderbouwing van besluiten over middelentoewijzing*

Een andere manier om extra kennis op te doen is door de netto-effecten te vergelijken van overeenkomende concrete acties die in verschillende gebieden zijn uitgevoerd. Als in de CIE's significante effectverschillen worden geconstateerd, duidt dat er hoogstwaarschijnlijk op dat de concrete acties qua kwaliteit aanmerkelijk uiteenliepen of dat sommige concrete acties beter bij de sociaaleconomische context pasten dan andere. Dankzij zulke vergelijkingen — doorgaans "meta-evaluaties" of "meta-analyses" genoemd — kunnen CIE-bevindingen verdergaand worden gegeneraliseerd. Ook leveren zij belangrijke toetsstenen voor beleidsmaatregelen (zoals opleiding, geïntegreerde diensten enz.)<sup>9</sup>. Om deze vergelijkingen op wetenschappelijk verantwoorde wijze te kunnen maken, is een afdoende aantal CIE's en een systematische uitvoering daarvan vereist. Goede coördinatie op nationaal of EU-niveau is derhalve een fundamentele voorwaarde om deze vergelijkingen frequenter te kunnen uitvoeren en ze van nut te doen zijn voor de beleidsvorming.

*Concrete acties in verschillende regio's of lidstaten met elkaar vergelijken*

<sup>9</sup> Zie bijvoorbeeld [Card D., Kluge J. en Weber A., 2017](#)



### 1.3. Waarom zijn contrafeitelijke effectonderzoeken in technisch opzicht een uitdaging?

Over een aantal methoden kan worden gezegd dat het “onbetrouwbare” pogingen zijn om interventie-effecten te ramen. Die methoden worden hier besproken om te laten zien dat CIE’s van nature complex zijn. Daarbij wordt niet verwezen naar de daadwerkelijke evaluatiepraktijk.

*Vorbij simplistische methoden*

Een beleidsmaker kan het effect van een opleidingsprogramma voor werklozen eerst willen evalueren door het inkomen van de cursisten na de opleiding te vergelijken met het inkomen van alle werklozen die niet aan het programma hebben deelgenomen. Vervolgens schrijft de beleidsmaker het geconstateerde inkomensverschil tussen deelnemers en niet-deelnemers toe aan het opleidingsprogramma.

Dit is geen geschikte strategie om het inkomenseffect van de opleiding te meten, want de niet-cursisten kunnen in een aantal belangrijke opzichten verschillen van de cursisten en die verschillen kunnen van invloed zijn op de resultaten. Met deze aanpak kunnen alternatieve verklaringen voor de inkomensverschillen die eventueel worden geconstateerd dus niet worden uitgesloten. Het kan bijvoorbeeld zo zijn dat de cursisten van zichzelf al over beter ontwikkelde vaardigheden beschikken dan de niet-cursisten. Anders geformuleerd: werklozen met beter ontwikkelde vaardigheden doen uit eigen beweging mee aan het opleidingsprogramma. Niet alleen het besluit tot deelname maar ook de resultaten worden dus door vaardigheden beïnvloed: werklozen met goed ontwikkelde vaardigheden kunnen vaker een hoger inkomen bedingen dan mensen met minder goed ontwikkelde vaardigheden. Het gevolg daarvan is dat inkomensverschillen die eventueel tussen de behandelingsgroep en de controlegroep worden waargenomen door het opleidingsprogramma en/of door inherente verschillen in vaardigheden kunnen zijn veroorzaakt, en dat het effect van de interventie dus wordt “overschat”.

Wanneer vaardigheden niet kunnen worden gemeten en wanneer, in geval van een effectraming waarbij de uitkomsten van de twee groepen met elkaar worden vergeleken, geen rekening kan worden gehouden met verschillen in inherente vaardigheden, wordt over het geschatte effect van het opleidingsprogramma wel gezegd dat dit effect aan **selectievertekening** lijdt. Om dit probleem te ondervangen, proberen beoordelaars zoveel mogelijk informatie te verzamelen over belangrijke factoren die van invloed zijn op de beslissing tot deelname en op de daaruit voortvloeiende resultaten. Die informatie wordt gebruikt om een geschikte controlegroep van niet-deelnemers samen te stellen — dus een groep die wat deze factoren betreft zo veel mogelijk lijkt op de behandelingsgroep — en om statistische analyses uit te voeren waarin rekening wordt gehouden met deze factoren. Daarbij beroepen beoordelaars zich vaak op de hypothese dat de selectie voor het programma op deze waarneembare factoren berust. Deze “selectiehypothese” kan in de regel niet eenvoudig worden gevalideerd en de beoordelaar moet aan de hand van institutionele factoren en de gedragstheorie aannemelijk kunnen maken dat het in de gegeven situatie om een plausibele hypothese gaat.

*Selectievertekening tegengaan*

Een tweede “onbetrouwbare” methode kan erin bestaan wanneer de beleidsmaker kijkt naar het inkomen van cursisten vóór en na de opleiding

*Voor en na de verandering*

en het verschil daartussen toeschrijft aan de opleidingsinterventie. Bij deze methode wordt in feite verondersteld dat het gemiddelde inkomen zonder de interventie gelijk blijft.

Nogmaals, tenzij op geloofwaardige gronden mag worden verondersteld dat er sprake is van stabiliteit in de tijd, is dit vrijwel nooit een geschikte strategie om het inkomenseffect van een opleiding te meten. Het is immers onvermijdelijk dat het inkomen van de cursisten na verloop van tijd verandert, door oorzaken die niets met de opleiding te maken hebben. Het komt bijvoorbeeld vaak voor dat het inkomen van cursisten vóór hun deelname terugloopt, ten dele wegens factoren van voorbijgaande aard. In veel gevallen is er dan sprake van inkomensherstel, met of zonder opleidingsinterventie<sup>10</sup>. De onbetrouwbare methode waarbij het opleidingseffect wordt gemeten aan de hand van het verschil tussen het inkomen onmiddellijk vóór en na programmadeelname gaat voorbij aan het feit dat het inkomen in veel gevallen hoe dan ook zou zijn gestegen.

Om dergelijke modellen te kunnen aanpassen, is een contrafeitelijke meting nodig, dat wil zeggen een meting die weergeeft hoe het inkomen van de cursisten zonder de opleidingsinterventie zou zijn veranderd. Dat contrafeitelijke resultaat kan bijvoorbeeld worden verkregen met behulp van een nauwkeurig gematchte controlegroep die de opleidingsinterventie niet ondergaat en waarvan het inkomen op dezelfde momenten wordt gepeild als bij de cursisten. Vervolgens wordt dan vaak de hypothese van de gemeenschappelijke trend gehanteerd, die zegt dat het inkomen van de cursisten en van de controlegroep zich zonder de interventie op dezelfde wijze zou hebben ontwikkeld.

De beperkingen van deze “onbetrouwbare” methoden zijn een aansporing om te zoeken naar evaluatiemethoden die plausibeler zijn. Zoals hierboven is aangestipt, kleven er aan het gebruik van plausibeler methoden echter grotere technische uitdagingen. Het volgende gedeelte van dit hoofdstuk bevat een kort overzicht van een aantal specifieke CIE-methoden die in de context van het ESF waarschijnlijk het meest relevant zijn.

#### 1.4. Overzicht van CIE-modellen en -methoden

Wanneer de aandacht vooral uitgaat naar de effecten van een interventie voor de deelnemers, worden de contrafeitelijke resultaten doorgaans geschat aan de hand van verzamelde gegevens over groepen niet-deelnemers die vergelijkbaar zijn met de mensen die wel deelnemen aan de onderzochte interventie. In tabel 1 aan het einde van dit hoofdstuk staat een beknopt overzicht van de methoden, enkele bijbehorende voordelen en beperkingen en de essentiële gegevenstypen die voor die methoden nodig zijn.

Het belangrijkste onderscheid tussen de diverse CIE's betreft de evaluatieopzet, die experimenteel of quasi-experimenteel kan zijn. De experimentele benadering wordt doorgaans een “gerandomiseerd onderzoek met een controlegroep” genoemd en soms ook wel een “sociaal experiment”.

---

<sup>10</sup> Dit patroon staat bekend als de “Ashenfelter Dip”, genoemd naar de econoom die dit als eerste besprak. Zie [Ashenfelter O., 1978](#)

De experimentele benadering wordt beschouwd als de gouden CIE-standaard voor het evalueren van de effecten van interventies die in een betrekkelijk kort tijdsbestek kunnen worden getoetst en aangepast. Deze benadering geldt in veel gevallen als ideaal. Een goed opgezet effectonderzoek is gericht op het verkrijgen van schattingen van niet-vertekende contrafeitelijke resultaten. In veel gevallen wordt aangenomen dat zulke niet-vertekende schattingen met een experimentele benadering kunnen worden verkregen. Het is vaak wenselijk om een bespreking van CIE-methoden te beginnen met een schets van de experimentele benadering. Dat heeft als reden dat met quasi-experimentele methoden in feite wordt getracht het ideaal van een experiment na te bootsen.

*Gerandomiseerde opzet: de gouden standaard*

Om de onderliggende principes te verduidelijken, zijn de hoofdkenmerken van de methoden in de bespreking van CIE-modellen zo eenvoudig mogelijk geformuleerd. In werkelijkheid kan de toepassing van deze methoden beduidend complexer zijn. Door problemen zoals uitval (*non-compliance*) — dat wil zeggen: personen die wel voor deelname aan een interventie waren geselecteerd maar niet hebben deelgenomen — kunnen de uitdagingen nog veel groter worden.

Er bestaat een breed scala aan methoden die in wezen zijn bedoeld als nabootsing van randomisatie. Deze methoden worden **quasi-experimenteel** genoemd. Binnen het bestek van deze leidraad is het ondoenlijk om al die methoden te bespreken of ze stuk voor stuk van een volledige en gedetailleerde technische beschrijving te voorzien. Globaal gezien worden in de context van het ESF de onderstaande quasi-experimentele methoden het vaakst toegepast: 1. propensity score matching; 2. difference-in-differences; 3. regressie-discontinuïteit; 4. instrumentele variabelen. In tabel 5 in hoofdstuk 3 staat een overzicht van de belangrijkste methoden en hun pluspunten. Om de kernbeginselen van elke methode goed te laten uitkomen, zijn de methoden sterk vereenvoudigd weergegeven. Bijlage I bevat aanvullende literatuur over quasi-experimentele methoden.

*Niet-gerandomiseerde of quasi-experimentele modellen*

## 1.5. Hoe kan een CIE in een breder evaluatiekader worden geïntegreerd?

Contrafeitelijke effectonderzoeken geven antwoord op bepaalde typen vragen met betrekking tot de causale effecten van interventies. Met deze methoden kan slechts tot op zekere hoogte een antwoord worden gegeven op andere vragen over een interventie. Het is nuttig om een onderscheid te maken tussen enerzijds evaluatievragen in verband met **causale verklaringen** en anderzijds evaluatievragen in verband met **causale beschrijvingen**. CIE's zijn bedoeld om de gevolgen van een interventie te **beschrijven**. Dergelijke methoden zijn minder geschikt om een **verklaring** te geven voor de mechanismen en contexten die de aanleiding voor causale verbanden vormen. Dit onderscheid is belangrijk omdat het licht werpt op de onderscheidende rol van CIE's<sup>11</sup>.

*Verklaring en beschrijving van causale verbanden*

Een goed geconcipeerd CIE vertelt beleidsmakers of een interventie de met die interventie beoogde gewijzigde resultaten heeft opgeleverd. Een CIE levert gegevens op over de omvang van de eventuele effecten, vertelt

*Wat een CIE beleidsmakers wel en niet kan vertellen*

<sup>11</sup>Zie Shadish W.R., Cook T.D., Campbell D.T., 2002, en [Stern E. et al., 2012](#)

de beleidsmaker of die effecten positief of negatief waren, maar biedt ook een mate van onzekerheid. Contrafeitelijke effectonderzoeken zijn minder geschikt om te verklaren waarom en hoe de met het CIE gemeten effecten zich hebben gemanifesteerd. Omgekeerd geldt dat het vaak moeilijk is om met een CIE vast te stellen waarom een interventie geen effect heeft gesorteerd, indien dat het geval blijkt te zijn.

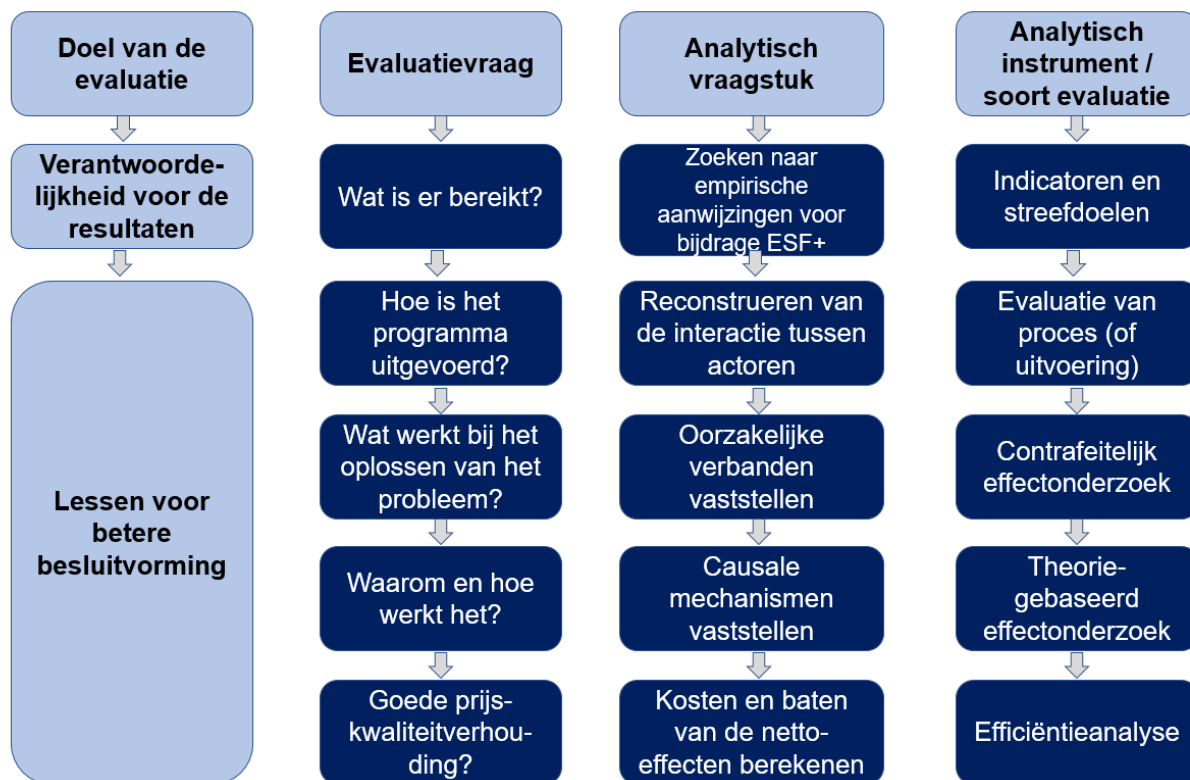
Voor de meeste beleidsorganen geldt dat belanghebbenden met causaal-beschrijvende vragen andere belangen en invalshoeken hebben dan belanghebbenden met causaal-verklarende vragen. Programmabeheerders en vakmensen houden zich meestal bezig met causaal-verklarende vragen. Mensen die middelen toewijzen en hooggeplaatste beleidsmakers die verantwoordelijk zijn voor het vaststellen van de begroting concentreren zich vaak op causaal-beschrijvende vragen. In de praktijk is het onderscheid tussen een causale verklaring en een causale beschrijving niet altijd helder. In sommige situaties kan met een CIE worden verklaard waarom bepaalde effecten werden geconstateerd, bijvoorbeeld door te onderzoeken welke effecten interventies op belangrijke subgroepen hadden. Toch is het cruciaal om zorgvuldig na te denken over het soort vragen dat belanghebbenden over een bepaalde interventie hebben en om de juiste benadering te kiezen voor de beantwoording van elk van die vragen. Als de hoofdvraag luidt of een interventie werkt, is een contrafeitelijk effectonderzoek in veel gevallen het aangewezen middel. Als de hoofdvraag luidt hoe een interventie werkt, dan zijn theoriegebaseerde methoden en procesevaluaties geschikter.

In de onderstaande figuur worden de verschillende soorten vragen en doelstellingen samengevat.

Uit deze bespreking blijkt dat CIE's binnen het kader van een evaluatieplan moeten worden opgesteld. Dat evaluatieplan moet bestaan uit verschillende evaluatievarianten waarmee uiteenlopende vragen van uiteenlopende belanghebbenden bij het beleid kunnen worden beantwoord. In de praktijk bevat een evaluatieplan zelden of nooit een CIE waarin een procesevaluatie ontbreekt.

Onder de noemer "evaluatie" wordt een breed scala methoden ingezet die voor allerlei verschillende doeleinden dienen. De hamvraag luidt hoe deze methoden op een nuttige manier kunnen worden gecombineerd, zodat beleidsleren wordt gestimuleerd. De echte uitdaging op dit gebied bestaat uit het vinden van de juiste combinatie van verschillende soorten evaluaties, die binnen de programmeringscyclus een verschillend doel dienen. Zoals reeds is besproken en in de volgende figuur wordt weergegeven, vullen CIE's, procesevaluaties en theoriegebaseerde methoden elkaar aan.

**Figuur 1** Verschillende evaluatietaken en -soorten



Bron: ontleend aan [Martini A., 2009](#)

Een gedegen evaluatiestrategie moet uit de volgende onderdelen bestaan:

- een theoriegebaseerd effectonderzoek;
- een procesevaluatie (of uitvoeringsevaluatie);
- een contrafeitelijk effectonderzoek (CIE);
- een efficiëntieanalyse.

*Soorten evaluaties combineren*

In deze leidraad worden alleen contrafeitelijke methoden voor effectonderzoeken grondig besproken. In de context van CIE's fungeren theoriegebaseerde methoden als middel om inzicht te krijgen in de bedoeling die achter de opzet van een interventie steekt.

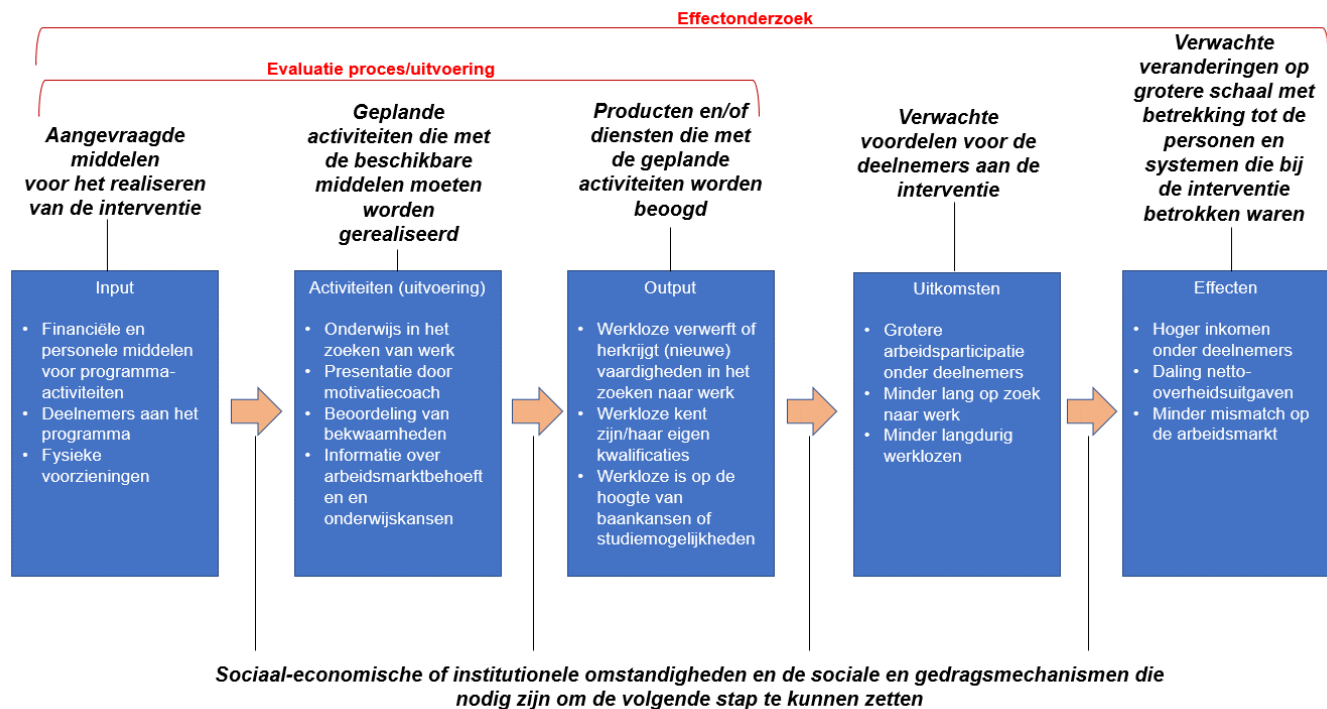
**Theoriegebaseerde evaluaties** worden in bepaalde gevallen gebruikt om de beoogde werking van de interventie te beschrijven en om te toetsen of de door de interventietheorie voorspelde verandering in resultaten zich heeft voorgedaan. Als zodanig kunnen theoriegebaseerde methoden voor effectbeoordelingen worden gebruikt, want zij bieden antwoord op de vraag "hoe" een effect is ontstaan. Ook kunnen zij worden gebruikt om interventie-effecten te onderzoeken in gevallen waarin geen CIE kan worden uitgevoerd. De volgende figuur toont een gestileerd voorbeeld van de interventielogica van een interventie die is gericht op het verbeteren van arbeidsbemiddelingsdiensten. Met een theoriegebaseerde evaluatie wordt onderzocht of de door de interventielogica gesuggereerde volgorde van de causale stappen, van input tot effect, empirisch standhoudt en onder welke voorwaarden en via welke maatschappelijke mechanismen dit mogelijk was. Een gedetailleerde beschrijving van het gebruik van theoriegebaseerde

*Met een theoriegebaseerde evaluatie wordt een interventielogica bedoeld*

methoden voor het vaststellen van effecten valt buiten het bestek van dit document.

Met een in het kader van een CIE uitgevoerde theoriegebaseerde evaluatie wordt nagegaan hoe een interventie is gepland en opgezet en hoe deze zou moeten functioneren. De essentie van de methode is om samen met de interventiebelanghebbenden te komen tot een gemeenschappelijke formulering van de logica waarop de interventie is gebaseerd, zoals in de volgende figuur vereenvoudigd is weergegeven. Elke interventie kent een programmalogica waarmee de input en activiteiten worden gekoppeld aan de outputs, de tussentijdse resultaten en later aan de resultaten op de langere termijn. Ook bij CIE's is het dus belangrijk om een interventielogica te formuleren: dankzij die logica is het makkelijker om een reconstructie te maken van het uitvoeringsproces en van de mogelijke invloed daarvan op de outputs en uitkomsten, en om de meest representatieve uitkomstvariabelen vast te stellen die in de contrafeitelijke analyse moeten worden getoetst.

**Figuur 2 Weergave van een logisch model of interventielogica**



Bron: Ontleend aan de [W.K. Kellogg Foundation, 2004](#) en [Bredgaard T., 2015](#)

Theoriegebaseerde evaluaties kunnen op diverse nuttige manieren aan contrafeitelijke effectonderzoeken worden gekoppeld. Een goed geformuleerde interventielogica kan dienen als aanknopingspunt voor de opzet van een CIE. Een goed omschreven interventielogica verschaft de ontwerper van een effectonderzoek onder meer informatie over de volgende aspecten: ...een aanvulling op CIE's

- Welke resultaten zijn belangrijk en moeten worden gemeten?
- Welk teken (positief of negatief) en welke omvang van interventie-effecten worden verwacht?
- Wie behoort tot de doelgroep en hoe kan er een controlegroep worden samengesteld?

- Hoelang kan het eventueel duren alvorens de effecten en resultaten van het programma zich manifesteren?
- Welke gegevens zijn er eventueel nodig om de deelname aan de interventie te kunnen meten?
- Hoe betrouwbaar is de controlegroep als maatstaf voor de contrafeitelijke situatie?

Het formuleren van een interventielogica kan ook helpen om eventuele ongewenste effecten op het spoor te komen, waarmee in de opzet van een CIE rekening kan worden gehouden. Een goed geformuleerde interventielogica kan de beoordelaar tot op zekere hoogte ook helpen bij het interpreteren van de resultaten van een contrafeitelijk effectonderzoek. Wat de interpretatie betreft, kan een **procesevaluatie** echter ook bijzonder leerzaam zijn.

*Procesevaluatie*

Een in het kader van een CIE uitgevoerde **procesevaluatie** dient een dubbel doel. Het eerste doel is om de “getrouwheid” te beoordelen; het tweede om het verschil te beoordelen tussen de ervaringen van de deelnemers aan de behandelingsgroep en aan de controlegroep.

Met de “getrouwheidsbeoordeling” wordt onderzocht in hoeverre de uitvoering van een interventie strookt met de opzet ervan. Met een procesevaluatie wordt nagegaan welke diensten daadwerkelijk aan de interventiedeelnemers ter beschikking zijn gesteld. Stemmen die overeen met datgene wat door de interventielogica werd beoogd? Wat verklaart de resultaatverschillen tussen locaties, voor zover er sprake is van dergelijke verschillen? De meeste interventies kennen zowel een management- als een effectlogica:

*Getrouwheidsbeoordeling*

- de managementlogica heeft betrekking op de wijze waarop uitvoerende organen geacht worden te reageren op de regels en prikkels van het programma;
- de effectlogica heeft betrekking op de wijze waarop de doelgroep van de interventie naar verwachting zal reageren, gelet op hetgeen daadwerkelijk is gerealiseerd.

Het getrouwheidsaspect van een procesanalyse levert informatie op over hetgeen daadwerkelijk met de interventie is bereikt en, bijgevolg, over de feitelijke oorzaken van de waargenomen effecten. Dit levert ook belangrijke feedback op voor het projectbeheer.

Vooraf bij contrafeitelijke effectonderzoeken is het belangrijk om de verschillen te onderzoeken. Zoals in deze leidraad grotendeels gebeurt, is het gebruikelijk om de effecten van de interventie centraal te stellen. Aan het effect op de resultaten gaat echter het effect op de inputs vooraf: het verschil in kansen tussen de behandelings- en de controlegroep dat met de interventie wordt gerealiseerd. In beginsel kan elk CIE “ondersteboven” worden gekeerd en kan de behandelingsgroep als controlegroep worden gebruikt om de resultaten te beoordelen voor de mensen die behoorden tot wat vóór de omkering de controlegroep werd genoemd. Dit betekent dat over de ervaringen van de controlegroep net zoveel bekend moet zijn als over de ervaringen van de behandelingsgroep, want de inputverschillen tussen de

*Verskil tussen behandelings- en controlegroep*

behandelingsgroep en de controlegroep worden in een CIE aangewezen als oorzaak van de verschillen in uitkomsten.

Terugkomend op het opleidingsprogramma is het mogelijk zich twee zeer verschillende uitgangssituaties voor te stellen. In de ene situatie wordt het opleidingsprogramma verzorgd in een generieke setting waarin niets vergelijkbaars voorhanden is. De leden van de controlegroep moeten het eenvoudigweg zonder opleiding stellen. Het kan echter ook zo zijn dat er wel een paar alternatieven bestaan. Er kan bijvoorbeeld een opleidingsaanbod zijn van bedrijven die zich op beroepsvoorbereiding hebben toegelegd. Als dat zo is, moet de procesanalyse voor zover mogelijk ook een beoordeling bevatten van het verschil in opleidingsdeelname tussen de behandelings- en de controlegroep. Er mag niet zonder meer worden aangenomen dat alle facetten van de behandeling buiten het bereik van de controlegroep liggen.

Procesevaluaties kunnen volledig losstaan van andere vormen van evaluatie, maar gezien het belang van procesevaluaties voor zowel het beheer als voor CIE's is het cruciaal dat zij samen met het effectonderzoek worden gepland.

*Geen CIE zonder procesevaluatie*

Goede procesanalyses kunnen bijdragen aan de getrouwheid en procesevaluaties voorzien interventies van een causale verklaring. Als de effectramingen eenmaal beschikbaar zijn, is het zonder goed opgezette procesevaluatie vaak lastig om een omvattende interpretatie aan de resultaten van een CIE te geven of om een taxatie te maken van de kosten die een kosten-batenanalyse met zich meebrengt.

Zoals hierboven is opgemerkt, kan een procesevaluatie ook behulpzaam zijn bij het interpreteren van de bevindingen van effectonderzoeken als wordt geschetst in welke context de interventie haar beslag kreeg. Begrip van de context is belangrijk omdat in die context de voorwaarden voor het succes van de interventie zijn gelegen en dit een idee geeft van de kans dat die interventie elders vergelijkbare effecten oplevert, in andere geografische gebieden of op andere momenten. Dit is met name van belang voor de discussie over de overdraagbaarheid van beleidsbenaderingen en om in het kader van leren en uitwisselingen over de grenzen heen de aandacht te vestigen op goede praktijken. Door een procesanalyse stijgt het vertrouwen in wat de **externe validiteit** van de evaluatieresultaten wordt genoemd.

Veelal bestaat de efficiëntieanalyse ofwel uit een beoordeling van de kosteneffectiviteit dan wel uit een volledige kosten-batenanalyse.

*Kosteneffectiviteit bepalen*

Bij een **analyse van de kosteneffectiviteit** worden de kosten van de interventie vergeleken met de effecten die in een CIE zijn vastgesteld. Simpel gezegd kan de kosteneffectiviteit worden berekend door het effect van een interventie (uitgedrukt in meeteenheden of standaardeenheden) te delen door de netto-uitvoeringskosten van de interventie per behandleenheid.

De kosteneffectiviteit van een opleidingsprogramma dat is bedoeld om werklozen aan een baan te helpen kan laten zien hoeveel geld er per deelnemer nodig is om hen weer op te nemen in het arbeidsproces.

*Baten en nettokosten vergelijken met een kosten-batenanalyse*



Voor degenen die verantwoordelijk zijn voor de toewijzing van middelen aan programma's is de kosteneffectiviteit een belangrijke maatstaf. Dankzij de kosten-batenverhoudingen van uiteenlopende interventies kunnen degenen die de middelen toewijzen zich een zeker oordeel vormen over de vraag welke interventies de beste prijs-kwaliteitverhouding hebben.

Met een **kosten-batenanalyse** worden de programma-effecten niet uitgedrukt in de desbetreffende meeteenheid of standaardeenheid, maar wordt geprobeerd om de met een CIE verkregen effectramingen uit te drukken in geld en die te vergelijken met de nettokosten van een interventie. Het doel van een kosten-batenanalyse is na te gaan of de in geld uitgedrukte baten van een programma hoger uitvallen dan de nettokosten ervan. Met een kosten-batenanalyse van een gemiddeld ESF-opleidingsprogramma worden de baten van de interventie voor de deelnemers, voor de overheid en voor de samenleving in bredere zin vergeleken met de nettokosten van de interventie. De programmabaten voor de deelnemers (doorgaans grotere inzetbaarheid op de arbeidsmarkt en een hoger inkomen) worden met behulp van een CIE bepaald. Om het nettovoordeel te berekenen worden die baten verminderd met de door de deelnemers afgedragen belastingen en andere baangerelateerde kosten. Vanuit de optiek van de overheid bestaan de baten van een interventie uit extra belastinginkomsten en minder sociale uitkeringen. De interventiekosten komen grotendeels voor haar rekening. De baten voor de samenleving als geheel worden berekend door de baten voor de deelnemers en voor de overheid bij elkaar op te tellen en die te verminderen met de totale kosten voor de deelnemers en de overheid.

Bij ESF-evaluaties zijn analyses van de kosteneffectiviteit en kosten-batenanalyses nog niet erg gangbaar. Toch zijn zulke analyses bijzonder nuttig wanneer moet worden besloten of voor een bepaalde interventie in de toekomst opnieuw geld wordt vrijgemaakt of om te achterhalen welke interventie uit een reeks vergelijkbare interventies het effectiefst is (zie het voorbeeld in het onderstaande tekstkader).

Effectramingen met behulp van een CIE vormen een essentieel onderdeel van zowel kosteneffectiviteitsanalyses als kosten-batenanalyses. In het eerste geval blijkt daaruit de mate van effectiviteit en in het tweede geval fungeren zij als essentieel instrument voor het schatten van de in geld uitgedrukte baten. Ook is duidelijk dat zowel voor onderzoeken naar de kosteneffectiviteit als voor kosten-batenanalyses nauwkeurige kostengegevens moeten worden verzameld aan de hand waarvan de nettokosten kunnen worden berekend. Deze activiteiten worden vaak als "kostenstudie" aangeduid. Bij sommige complexe evaluaties op basis van gemengde methoden worden kostenstudies vaak in de procesevaluatie geïntegreerd. Daarbij is het mogelijk de onderzoeksinstrumenten zo aan te passen dat belangrijke kostengegevens kunnen worden verzameld.

### Tekstkader 1 Voorbeeld van een CIE voor het vergelijken van kosten en baten

In [Bazzoli M. et al., 2018](#) staat een voorbeeld van de wijze waarop met een contrafeitelijke methode de kosten en baten worden geschat. In dit onderzoek staan programma's voor beroepsopleidingen centraal die in de jaren 2010-2011 in de Italiaanse Autonome Provincie Trento zijn uitgevoerd en in het kader waarvan meer dan 300 uur aan opleidingsactiviteiten werd aangeboden. Er werden twee categorieën cursussen geëvalueerd: cursussen die met provinciale middelen werden gefinancierd en cursussen die uit het Europees Sociaal Fonds werden gefinancierd. De cursussen telden respectievelijk 954 en 205 deelnemers.

Dit waren de belangrijkste stappen voor de uitvoering van de kosten-batenanalyse: 1) de auteurs beoordeelden wat de cursussen hadden betekend voor de kans dat de deelnemers binnen drie jaar na het begin van de cursus een baan hadden gevonden. Daarvoor maakten zij gebruik van propensity score matching<sup>12</sup>; 2) er werd een schatting gemaakt van het effect van de cursussen op het bruto-inkomen tot en met eind 2013; 3) de auteurs maakten een schatting van het bedrag aan extra belastingopbrengsten (op basis van het inkomenseffect) en van het bedrag aan overheidsmiddelen dat dankzij de afname van het aantal mensen met een werkloosheidsuitkering werd bespaard; 4) de kosten en de baten van de cursussen werden tegen elkaar afgezet.

Er werd gebruikgemaakt van diverse administratieve datasets: a) monitoringgegevens over de deelnemers en hun kenmerken; b) gegevens uit registers van openbare diensten voor arbeidsvoorziening (*Centri per l'impiego* of arbeidsbureaus) over de werklozen om de controlegroep te kunnen bepalen; c) gegevens uit de COB-databank, het archief met verplichte kennisgeving van arbeidsovereenkomsten door bedrijven aan de openbare diensten voor arbeidsvoorziening (ODA's), om de arbeidsstatus van de deelnemers aan de behandelings- en controlegroep vast te stellen, zowel voor als na deelname aan de opleiding; d) gegevens uit archieven van de belastingdienst om het inkomen van de personen in kwestie te berekenen en e) gegevens van het Italiaanse nationale instituut voor de sociale zekerheid (INPS) voor informatie over door de personen in kwestie ontvangen werkloosheidsuitkeringen.

Na 36 maanden lag de kans op het hebben van een baan in de behandelingsgroep die had deelgenomen aan de door de provincie gefinancierde cursussen ongeveer 5 procentpunten hoger dan in de controlegroep, terwijl het effect van de uit het ESF gefinancierde cursussen met ongeveer 28 procentpunten veel groter was. In de drie jaar volgend op de interventie verdienden de mensen die aan de door de provincie gefinancierde cursussen hadden deelgenomen gemiddeld 2 250 EUR per jaar meer dan de controlegroep, terwijl de deelnemers aan de uit het ESF gefinancierde cursussen jaarlijks 4 106 EUR meer verdienden dan de controlegroep. Op basis van de gegevens konden de auteurs ook een raming maken van de baten voor de overheid, in de vorm van hogere belastinginkomsten en lagere uitgaven aan sociale uitkeringen. In elk onderzocht jaar betaalden de door de provincie gefinancierde cursisten 126 EUR meer inkomstenbelasting dan de controlegroep. De schatting voor de uit het ESF gefinancierde cursisten kwam uit op 318 EUR per jaar. Wat de daling van de door de overheid betaalde werkloosheidsuitkeringen betreft, was de betekenis van de cursussen verwaarloosbaar, hoogstwaarschijnlijk omdat de hoogte van die uitkering samenhangt met de duur van het arbeidsverleden op het moment waarop iemand werkloos wordt en het veelal jonge deelnemers met geringe werkervaring betrof.

De kosten van de interventies bedroegen per deelnemer 4 800 EUR bij de door de provincie gefinancierde cursussen en 14 500 EUR bij de uit het ESF gefinancierde cursussen. Daarnaast vergeleken de auteurs voor beide typen cursussen de kosten en baten op individueel niveau. Gemiddeld genomen bleek dat de kosten in de periode 2010-2013 groter waren dan de baten<sup>13</sup>.

<sup>12</sup> Om precies te zijn pasten de auteurs de BRA-schattingmethode (*blocking with regression adjustment*) toe.

<sup>13</sup> Een "vergelijkbare" exercitie staat in [Lammers M. en Kok L., 2021](#)

## Hoofdstuk 2. Praktische overwegingen bij het opzetten van een CIE

In dit hoofdstuk wordt ingegaan op praktische kwesties waarop bij de voorbereiding van een evaluatie moet worden gelet. Dit hoofdstuk moet worden gebruikt bij het plannen van evaluatie-activiteiten, bij het selecteren van interventies die aan een CIE-methode worden onderworpen en bij het vaststellen van de hoofdvragen die bij het opstellen van een CIE moeten worden beantwoord.

*Een CIE  
voorbereiden*

Als vertrekpunt geldt de veronderstelling dat een programmabeheerder van een beheerautoriteit (of een beheerder van een intermediaire instantie die verantwoordelijk is voor de uitvoering van een ESF-interventie) zich beraadt over de vraag welke interventies moeten worden geëvalueerd en welke strategieën geschikt zijn voor het integreren van een CIE. Verder wordt verondersteld dat de ambtenaren van een beheerautoriteit de evaluaties niet zelf uitvoeren, maar dat die worden uitbesteed of dat er evaluatiediensten worden betrokken van externe deskundigen. Hoewel het CIE door een contractant zal worden uitgevoerd, moet het effectonderzoek vóór de opdrachtverlening worden gepland en voorbereid door de beheerautoriteit (of intermediaire instantie).

De evaluatiestrategie moet in het evaluatieplan worden vastgelegd, wat ook geldt voor de verschillende soorten evaluaties die in het vorige hoofdstuk zijn beschreven.

Evaluatieplannen zijn voor alle programma's verplicht en moeten uiterlijk een jaar na goedkeuring van het programma door het toezichtcomité zijn goedgekeurd<sup>14</sup>. De plannen moeten aan het begin van de programmeringsperiode worden opgesteld en het volgende bevatten: de regelingen voor het evaluatieproces (de governance van de evaluaties en het verband tussen evaluatie en toezicht), de daadwerkelijke evaluatieactiviteiten (een indicatieve lijst van uit te voeren evaluaties, de reikwijdte van elke evaluatie, de hoofdvragen, de benodigde gegevens, het mogelijke gebruik, een indicatief tijdschema, de beheersstructuur enz.), het tijdschema van de evaluaties, de totale begroting en de opbouw van evaluatiecapaciteit.

Evaluatieplannen hebben doorgaans een algemeen karakter, terwijl de planning van een CIE juist om een minutieuze voorbereiding vraagt. Dit gebeurt idealiter tijdens het opstellen van het evaluatieplan. Bepaalde details volgen eventueel in een later stadium. Beheerautoriteiten en intermediaire instanties moeten zich echter realiseren dat het meestal niet eenvoudig is om contacten met belanghebbenden aan te knopen en andere regelingen te treffen die nodig zijn om gegevens over interventies te kunnen inwinnen en dat dit ruim van tevoren moet worden gepland.

---

<sup>14</sup> Evaluatieplannen zijn verplicht op grond van artikel 44, leden 5 en 6, van Verordening (EU) 2021/1060 van het Europees Parlement en de Raad. Alle programma's moeten in een evaluatieplan worden opgenomen, maar één plan kan de voorgenomen evaluaties van meerdere programma's omvatten.

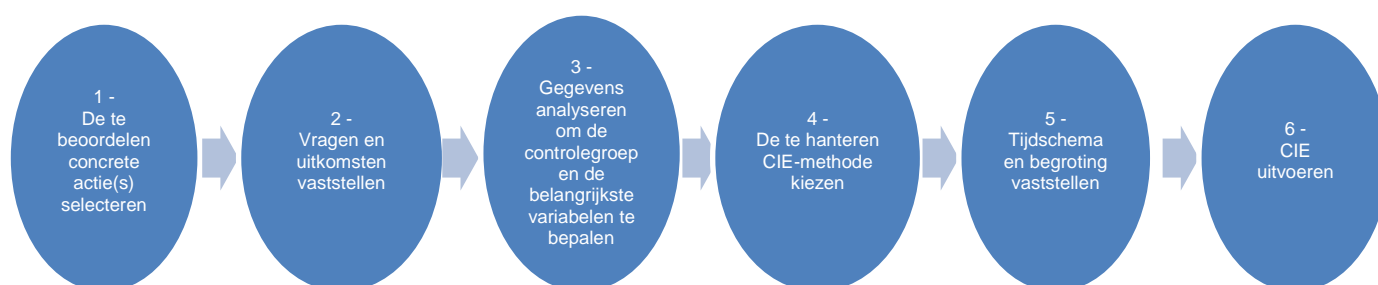
Deze leidraad spitst zich toe op methoden voor het opzetten van een evaluatieschema voor welbepaalde interventies die in aanmerking komen voor een CIE. Dit schema moet een onderdeel van het evaluatieplan zijn of een operationele stap vormen die aansluit op een evaluatieplan.

*Een evaluatieschema ontwikkelen voor specifieke interventies*

Niet alle uit het ESF gefinancierde interventies kunnen aan een contrafeitelijk effectonderzoek worden onderworpen. Beleidsmakers moeten kiezen waar ze zich op willen concentreren en er moet een procedure zijn om te bepalen naar welke interventies een effectonderzoek zal worden uitgevoerd. In deze leidraad wordt een aantal aspecten aangestipt waarop de beheerautoriteiten moeten letten bij het kiezen van geschikte interventies. Daarnaast is de kerndoelstelling van deze leidraad om de verantwoordelijken voor de opdrachtverlening van CIE's te laten reflecteren op enkele uitdagingen bij het opzetten van een geslaagd effectonderzoek en hen op die manier te helpen bij het opzetten van evaluatieschema's voor de diverse CIE's die zij in overweging hebben.

In deze leidraad wordt aangenomen dat, nadat is besloten naar welke interventies een CIE zal worden uitgevoerd, de beheerautoriteiten voor alle geselecteerde interventies een evaluatieschema moeten opstellen. In deze leidraad wordt het begrip "schema" gebruikt om deze activiteit te onderscheiden van de krachtens Verordening (EU) 2021/1060 (gemeenschappelijke bepalingen) verplichte formele "evaluatieplannen" voor de programmeringsperiode 2021-2027. Meer in het bijzonder duidt de term "schema" op een reeks gestandaardiseerde activiteiten die nodig is voor het definiëren en uitvoeren van een CIE. Zoals in de onderstaande figuur wordt aangegeven, moeten die activiteiten voor aanvang van een CIE worden voorbereid.

**Figuur 3 Hoofdvolgorde van CIE-activiteiten**



Deze schema's vormen het fundament voor de opdrachtverlening van CIE's en het uitgangspunt op basis waarvan contractanten een grondig en waardevol onderzoek kunnen uitvoeren. In de rest van dit hoofdstuk wordt ingegaan op de vragen die in de evaluatieplanning aan de orde moeten komen. De evaluatieschema's moeten worden toegespitst op de concrete omstandigheden waaronder de interventie wordt uitgevoerd. Hoe die concrete omstandigheden eruitzien, laat zich onmogelijk voorspellen. Daarom wordt in deze leidraad stilgestaan bij vragen die a) in de schema's aan de orde moeten komen; of b) die uitnodigen tot bezinning over de uitdagingen waarop die schema's een antwoord moeten bieden.

*Vragen die in de evaluatieplanning moeten worden beantwoord*

Nadat eerst wordt ingegaan op een aantal kwesties die aandacht behoeven als wordt nagedacht over de vraag welke interventies eventueel aan een CIE worden onderworpen, en of het met de beschikbare soorten gegevens überhaupt mogelijk is om een CIE uit te voeren, richt de aandacht zich op enkele belangrijke vragen die bij het opzetten van een evaluatieschema moeten worden bestudeerd. Het is noodzakelijk dat dit op schrift wordt gesteld voordat opdracht tot een CIE — of een breder evaluatieonderzoek — wordt gegeven, zodat de opdracht kan worden geformuleerd en een contractant kan worden aangesteld. In de tabel hieronder staan de belangrijkste inhoudelijke punten van een dergelijk evaluatieschema.

**Tabel 1 Aanbevolen inhoud van een evaluatieschema**

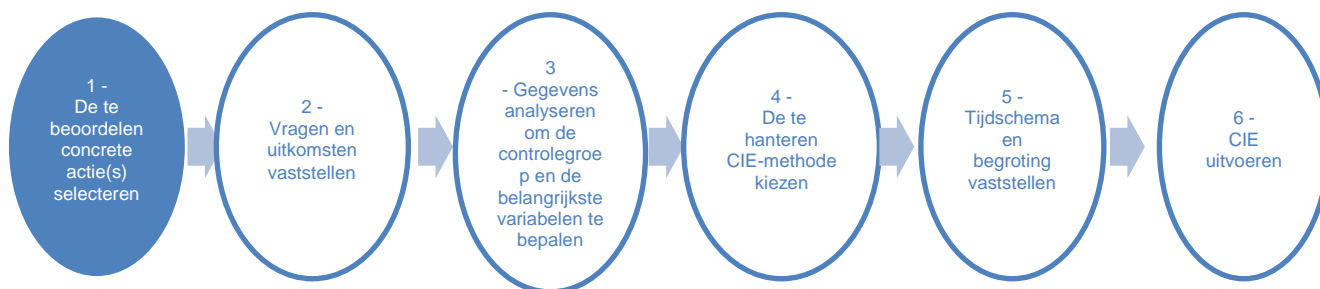
Belangrijkste stappen bij het voorbereiden en uitvoeren van een CIE	Inhoud
<b>1. De te beoordelen concrete actie(s) selecteren</b>	<ul style="list-style-type: none"> <li>– de concrete acties of de soorten concrete acties (waar mogelijk behorend tot een vooraf bepaalde typologie) van het ESF die met behulp van een CIE moeten worden geëvalueerd;</li> <li>– beknopte beschrijving van de werklogica van de geselecteerde concrete acties (doelstellingen, belangrijkste selectiecriteria en doelpopulatie, vormen van steun, uitvoeringsmodel, geschatte start- en einddata, indicatieve begroting).</li> </ul>
<b>2. De evaluatievragen en uitkomstvariabelen vaststellen</b>	<ul style="list-style-type: none"> <li>– evaluatievragen;</li> <li>– het verwachte functioneren van de concrete acties (“interventiologica”) en de uitkomstvariabelen (zoals arbeidsstatus, veranderingen in inkomen, armoedestatus, gemiddelde examenscores enz.).</li> </ul>
<b>3. Gegevens analyseren om de controlegroep te bepalen en de uitkomstvariabelen te meten</b>	<ul style="list-style-type: none"> <li>– de te gebruiken administratieve gegevens (bv. werklozenregisters, belastingregisters, administratieve gegevens met betrekking tot verzekeringen, scholenregisters of databanken met studenten ...) of andere gegevens (bv. enquêtes, big data);</li> <li>- variabelen van de databank of enquête die voor het CIE worden gebruikt;</li> <li>- tijdreeksen van de benodigde variabelen;</li> <li>- belangrijkste voorschriften en problemen omtrent gegevenstoegang (rechtstreekse toegankelijkheid voor de beheerautoriteit, instemming van andere overheden nodig, regelgeving en beperkingen op het gebied van privacy).</li> </ul>
<b>4. De CIE-methode kiezen</b>	<ul style="list-style-type: none"> <li>– CIE-methode waarvoor eventueel wordt gekozen (deze kan later worden uitgewerkt, maar ten behoeve van een samenhangend evaluatieproces moet van tevoren worden aangegeven of gebruik zal worden gemaakt van gerandomiseerd onderzoek met een controlegroep of andere methoden).</li> </ul>
<b>5. Het tijdschema en de begroting vaststellen</b>	<ul style="list-style-type: none"> <li>– tijdschema van de CIE's, belangrijkste mijlpalen: a) besluit over een experimentele of quasi-experimentele aanpak; b) uitvoerige evaluatievragen; c) opdracht formuleren; d) beoordelaar kiezen; e) gegevensvoorbereiding; f) gegevensanalyse; g) tussentijdse verslagen en eindverslagen; h) resultaten valideren; i) resultaten en lessen verspreiden;</li> <li>– de voor het CIE beschikbare begroting.</li> </ul>
<b>6. De evaluatie uitvoeren</b>	<ul style="list-style-type: none"> <li>– beoordelaar kiezen;</li> <li>– toezicht op de uitvoering van het CIE;</li> <li>– verslaglegging over het CIE;</li> <li>– verspreiding van de resultaten (voornaamste belanghebbenden die moeten worden ingeschakeld, belangrijkste instrumenten).</li> </ul>

## Tekstkader 2 Integratie van CIE's in een breder kader

Veel van de CIE's die in de lidstaten naar uit het ESF gefinancierde interventies worden uitgevoerd, zijn geïntegreerd in een breder evaluatiekader:

- bij het operationele programma “ESF Bund Deutschland 2014-2020” in Duitsland maakt het contrafeitelijke effectonderzoek naar het programma voor de arbeidsmarktintegratie van langdurig werklozen, dat uit hoofde van investeringsprioriteit IP 9.i werd gefinancierd, deel uit van een bredere evaluatiestrategie die voorziet in jaarlijkse tussentijdse verslagen waarin ook onderwerpen aan bod komen die te maken hebben met de uitvoering van de interventies. In de periode 2017-2021 werden vier verslagen opgesteld<sup>15</sup>. Hetzelfde geldt voor de evaluatie van de ESF-maatregelen ter ondersteuning van de integratie van langdurig werklozen in Baden-Württemberg. In dat geval had de contrafeitelijke analyse een meer kwalitatieve insteek waarbij de beoordeling van de interventies door de deelnemers centraal stond<sup>16</sup>.
- In Piëmont (Italië) bestond één stap in een uitgebreidere lopende evaluatie uit een contrafeitelijk effectonderzoek waarin de werkgelegenheidseffecten van vouchers voor begeleiding van kwetsbare mensen werden beoordeeld, gefinancierd in het kader van investeringsprioriteiten 8.i en 9.i. In twee initiële verslagen werden uitvoeringsproblemen onderzocht en werd gekeken naar de subjectieve indrukken van de deelnemers (april 2018 en februari 2019), terwijl in twee vervolgverslagen, van juli 2019 en eind 2020, de nadruk lag op de werkgelegenheidseffecten en een contrafeitelijke methode werd gebruikt<sup>17</sup>.
- Het CIE waarmee in Marche (Italië) in 2020 werd gekeken naar de effecten van op langdurig werklozen gerichte interventies was een effectanalyse met een thematische focus. Deze volgde op een meer algemene effectanalyse uit 2019 naar concrete acties van het ESF die op werklozen waren gericht (re-integratieverslag)<sup>18</sup>.
- In Polen werd in het kader van het operationele programma “Kennis Onderwijs Groei 2014-2020” tussen eind 2015 en mei 2020 opdracht gegeven tot een aantal evaluaties (acht verslagen<sup>19</sup>) om de ESF- en YEI-steun aan jongeren vanuit diverse invalshoeken en met verschillende analytische methoden te onderzoeken, zoals kwalitatieve analyses in de vorm van enquêtes en interviews, macromodellen, activiteiten te velde en een contrafeitelijke benadering).

### 2.1. De interventies voor een effectonderzoek selecteren



<sup>15</sup> Zie [Boockmann B. et al., 2017](#) – [Boockmann B. et al., 2018](#) – [Boockmann B. et al., 2019](#) en [Boockmann B. et al., 2021](#). In 2019 en 2021 werd een CIE toegepast.

<sup>16</sup> Zie [Hunger K. en Sattler K., 2017](#) en [Scheller F. en Seidel K., 2020](#). In 2020 werd een CIE toegepast.

<sup>17</sup> Zie [Pomatto G., 2017](#) – [Pomatto G., 2019](#) – [Poy S., 2019](#); [Poy S., 2020](#). In 2019 en 2020 werd een CIE toegepast.

<sup>18</sup> Zie [Pompili M., Giorgetti I., 2020](#) – [Pompili M., Giorgetti I., 2020a](#)

<sup>19</sup> Zie [Instytut Badań Strukturalnych - Imapp - IQS, 2015](#) – [Baran J. et al., 2016](#) – [Baran J. et al., 2017](#) – [Baran J. et al., 2018](#) – [Baran J. et al., 2018a](#) – [Palczyńska M. et al., 2019](#) – [Kalinowski H., 2020](#) – [Kalinowski H. et al., 2020](#). In 2017 en 2020 werd een CIE toegepast.

Voor het selecteren van interventies voor effectonderzoeken zijn drie stappen nodig:

*Selectiecriteria  
voor interventies*

1. Er moeten strategische thema's worden vastgesteld.
2. Wanneer duidelijk is wat de strategische prioriteiten zijn, moet worden getoetst of de afzonderlijke interventies voldoen aan de basisvereisten voor de contrafeitelijke methode en in hoeverre zij innovatief zijn en/of een significante bijdrage aan de kennisbasis kunnen leveren.
3. Er moet duidelijkheid bestaan omtrent de beschikbaarheid of potentiële beschikbaarheid van de soorten gegevens die nodig zijn om een CIE te kunnen uitvoeren. Dit derde punt is tot op heden een belangrijke belemmering gebleken voor de uitvoering van contrafeitelijke effectonderzoeken naar ESF-interventies en verdient bijzondere aandacht.

### Tekstkader 3 Vragen bij het selecteren van interventies voor een CIE

Niet alle interventies lenen zich voor een CIE en het is over het algemeen niet kosteneffectief om voor alle interventies een CIE uit te voeren. De beheerautoriteiten moeten een manier voor de toewijzing van middelen bedenken die de grootst mogelijke baten oplevert. Deze keuzes worden geacht door te klinken in het evaluatieplan. Bij de planning van de CIE's moet op drie belangrijke aspecten worden gelet: a) strategische prioriteiten; b) de haalbaarheid van een CIE; en c) de beschikbaarheid van de benodigde gegevens.

De omvang, de beleidsontwikkeling, de punten van onzekerheid en de behoefte aan kennis zijn van invloed op de **evaluatiestrategie**. De beheerautoriteiten moeten de volgende vragen stellen:

- Is een rechtvaardiging van de uitgaven bijzonder belangrijk, gelet op de grote bedragen die voor deze interventie zijn uitgetrokken? Het is vrij eenvoudig te achterhalen om welke interventies het gaat. Deze interventies ontvangen immers het leeuwendeel van de middelen die per specifieke doelstelling zijn toegewezen (zoals gedefinieerd in artikel 2 van Verordening (EU) 2021/1057 tot oprichting van het ESF+).
- Vormt de maatregel de spil van een hervormingsproces en zullen de resultaten van de evaluatie naar verwachting bijdragen aan een kritische beoordeling van de inspanning? Deze interventies houden verband met recente hervormingen van het arbeidsmarktbeleid, het onderwijsbeleid of het sociaal beleid. Hoewel voor die interventies wellicht geen grote bedragen zijn uitgetrokken, zijn zij wel cruciaal voor het welslagen van de hervorming.
- Is de interventie innovatief en wordt deze, alvorens te worden opgeschaald, getest met behulp van een proefproject? Voor deze interventies zijn wellicht geen enorme bedragen uitgetrokken, maar een vroegtijdige beoordeling ervan is noodzakelijk om te kunnen besluiten of het experiment een vervolg krijgt en wordt uitgebreid dan wel wordt stopgezet.
- Is de interventie gericht op gebieden waarvoor aanvullend bewijs van effectiviteit nodig is? In deze categorie vallen interventies die in het verleden niet zijn geëvalueerd of waarvan de laatste evaluatie zo oud is dat een actualisering nodig is. Omdat CIE's nog maar sinds kort op grote schaal worden toegepast en de "netto-effecten" onbekend zijn, komt dit type interventie binnen de ESF+-programma's volop voor.

**Haalbaarheid** heeft zowel betrekking op de kenmerken van de interventies als op de omstandigheden waaronder zij worden geïntroduceerd. Planners moeten op de volgende vragen een bevestigend antwoord kunnen geven:

- Is de behandeling in het kader van de interventie afzonderlijk, onderscheidend en voldoende homogeen?
- Is de vergelijking tussen de behandelings- en de controlegroep zinvol genoeg om het effect te kunnen

meten?

- Is er sprake van een grote en goed afgebakende doelpopulatie?
- Vertoont de theorie die het verband vormt tussen de interventie en de beoogde uitkomsten logische samenhang?
- Laat de behandelingsgroep zich binnen de doelpopulatie duidelijk aanwijzen?
- Volstaat de omvang van de behandelingsgroep?
- Is het mogelijk om een geloofwaardige controlegroep aan te wijzen?
- Kan het verschil tussen de ervaringen van de behandelingsgroep en van de controlegroep lang genoeg in stand worden gehouden om de effecten te kunnen meten?

**Gegevens** spelen een essentiële rol. Een CIE draait om “meten” en voor metingen zijn kwantitatieve gegevens nodig over de behandelings- en de controlegroep en over de context waarin de evaluatie wordt uitgevoerd. Welke gegevens precies nodig zijn, is meestal afhankelijk van de interventielogica en van de strategie die voor het bepalen van de contrafeitelijkheid wordt gevolgd. De beheerautoriteit die een CIE plant, moet zich bij de selectie van interventies voor een CIE deze vragen stellen:

- Welke informatie over de leden van de doelgroep en de controlegroep is onmisbaar?
- Welke informatie over de aard van de interventie zoals die daadwerkelijk op de behandelingsgroep wordt toegepast is onmisbaar?
- Ondergaat de controlegroep geen behandeling of een andere behandeling? Zijn daar gegevens over?
- Welke gegevens zijn er in administratieve of andere bronnen beschikbaar?
- Zijn er gegevens beschikbaar die een beschrijving geven van individuele loopbanen?
- Kunnen geïndividualiseerde gegevens uit verschillende bronnen aan elkaar worden gekoppeld?

Verderop in dit hoofdstuk volgt nadere informatie over deze onderwerpen.

### **2.1.1. Interventies voor effectonderzoeken prioriteren**

Voordat met het oog op een CIE prioriteit wordt toegekend aan concrete interventies, moet worden nagedacht over bredere strategische thema's. De nadruk moet liggen op het selecteren van interventies waarvoor geldt dat een effectonderzoek daarnaar het grootste rendement oplevert qua kennis over de vraag wat er werkt. Van een goed opgezette, grondige evaluatie profiteren niet alleen de beheerautoriteiten en de lidstaten die de opdracht hebben verleend, maar ook andere lidstaten en hun beheerautoriteiten, andere belanghebbenden en de Commissie.

#### **Rol bij het rechtvaardigen van uitgaven**

Gelet op het feit dat bij CIE's de nadruk ligt op de beantwoording van vragen die van wezenlijk belang zijn voor beleidsmakers, en met name voor degenen die verantwoordelijk zijn voor besluiten inzake de toewijzing van middelen, is het logisch dat effectonderzoeken worden gericht op programma's en interventies die bijzonder veel middelen vergen. Naarmate met een bepaald programma of een bepaalde interventie meer tijd of andere middelen gemoeid zijn, wordt het belangrijker om te weten of dat programma of die interventie werkt en dus of de gegenereerde baten de gemaakte kosten overtreffen. Kostbare interventies die geen maatschappelijke of economische waarde genereren, moeten eventueel opnieuw worden bezien, terwijl interventies waarvan is bewezen dat die meerwaarde opleveren wellicht extra financiering en aandacht verdienen.

*Nadruk op interventies die veel middelen vergen*



Uit de resultaten van recente evaluaties van ESF-interventies die in de programmeringsperiode 2014-2020 zijn gefinancierd, is gebleken dat er behoefte bestaat aan strategische aanpassingen en grotere aandacht voor de belangrijkste beleidsdoelstellingen. Interventies op het gebied van werkgelegenheid en arbeidsmobiliteit bleken minder effectief voor ouderen en voor mensen met grotere afstand tot de arbeidsmarkt. Voor deze groepen zijn nieuwe en effectievere instrumenten nodig<sup>20</sup>. Hoewel er bij sociale beleidsinterventies sprake is van grote verschillen in de kosten per deelnemer en per type concrete actie, blijft een systematische kosten-batenanalyse over het algemeen achterwege<sup>21</sup>. Interventies op het gebied van onderwijs en opleiding hebben wel positieve resultaten laten zien, maar vanwege het geringe aantal effectonderzoeken is het niet mogelijk om een allesomvattend oordeel te geven over de langetermijneffecten ervan<sup>22</sup>. Met CIE's kunnen interventies onder de loep worden genomen met als doel te achterhalen welke methoden voor bepaalde doelgroepen het effectiefst en efficiëntst zijn, waardoor de "prijs-kwaliteitverhouding" van de nieuwe ESF+-programma's wordt gemaximaliseerd.

### **Bijdrage van een interventie aan een hervormingsproces**

Interventies die als hoeksteen dienen van een breder hervormingsprogramma trekken vaak aanzienlijke financiering aan. Wanneer een ESF-interventie van essentieel belang is voor een strategie inzake sociale inclusie of een cruciaal aspect vormt van een programma voor arbeidsmarktactivering, genereert dat uiteraard extra aandacht voor die interventie.

*Interventies die  
bijdragen aan  
beleidsinnovatie*

### **Innovatief en verkennend**

Het ligt voor de hand dat nieuwe en innovatieve proefinterventies in aanmerking komen voor een CIE. Om de effecten van interventies met behulp van een proefproject te kunnen toetsen, is ontegenzeggelijk een grondige evaluatie vereist. Een evaluatie in de vorm van een goed geconcipeerd CIE is des te belangrijker wanneer er een duidelijke toezegging ligt om de interventie, als die geslaagd wordt bevonden, op te schalen of breder uit te rollen.

### **Bijdrage aan kennisopbouw**

Er is extra reden om aandacht en middelen op bepaalde programmagebieden — en op bepaalde interventies op die gebieden — te concentreren wanneer er weinig tot geen aanwijzingen zijn over wat op het betreffende beleidsterrein wel of niet werkt. Dat kan bijvoorbeeld het geval zijn wanneer er reële onzekerheid heerst omtrent het toekomstige beleid en het gevaar bestaat dat te zwaar wordt geleund op gegevens die mogelijk niet rechtstreeks relevant zijn (bv. gegevens uit andere landen).

*Empirische  
gegevens  
genereren*

Kwalitatief hoogwaardige evaluaties kunnen als een publiek goed worden beschouwd. De baten die zij qua kennis opleveren vallen ook toe aan belanghebbenden die niet tot een specifieke beheerautoriteit behoren. Daarom is het belangrijk om na te gaan welke belanghebbenden van het voorgestelde effectonderzoek zouden kunnen profiteren. Dat kunnen intermediaire instanties zijn of agentschappen die zich binnen hetzelfde programma met interventies bezighouden, andere beheerautoriteiten of

<sup>20</sup> [Fondazione G. Brodolini, Metis GmbH, Applica, Ockham IPS \(2020\)](#)

<sup>21</sup> [ICF, Cambridge Econometrics and Eurocentre \(2020\)](#)

<sup>22</sup> [Ecorys, Ismeri Europa \(2020\)](#)

intermediaire instanties in de betrokken lidstaat, of agentschappen en instellingen die zich met nationale of regionale fondsen bezighouden. De Europese Commissie is een andere voor de hand liggende externe belanghebbende waaraan moet worden gedacht. En ook in andere lidstaten zijn er belanghebbenden die lessen uit een evaluatie kunnen trekken. Beleidsmakers en programmabeheerders kunnen een belangrijke bijdrage leveren aan wederzijds leren door niet alleen rekening te houden met de behoeften van de direct betrokken belanghebbenden, maar ook met die van belanghebbenden die op wat grotere afstand staan.

Een laatste strategische overweging bij het kiezen van aandachtsgebieden tijdens het opzetten van CIE's is om te kijken naar interventies die het nut van CIE's over het voetlicht kunnen brengen en als model kunnen fungeren.

*Pleidooi voor CIE-methoden*

### **2.1.2. Interventies selecteren die geschikt zijn voor een contrafeitelijke aanpak**

Na te hebben stilgestaan bij bredere strategische kwesties die aanleiding kunnen zijn om bepaalde interventies voor een CIE te selecteren, wordt in dit deel ingegaan op de specifieke kenmerken van interventies waardoor zij mogelijk geschikt zijn voor een contrafeitelijke aanpak. Die kenmerken kunnen talrijk en gevarieerd zijn. Sommige interventiekenmerken lenen zich onder bepaalde omstandigheden voor een CIE, maar belemmeren in andere omstandigheden de uitvoering daarvan. Als gevolg daarvan kan geen omvattende lijst overwegingen worden gegeven. Wel loont het om de interventiekenmerken te noemen die vaker een geslaagd CIE opleveren.

#### **Tekstkader 4 Soorten ESF-interventies en -doelgroepen die het vaakst worden gekozen voor contrafeitelijke effectonderzoeken**

De meeste CIE's naar ESF-interventies hebben betrekking op actief arbeidsmarktbeleid dat is gericht op werklozen en subgroepen werklozen die met een specifiek vorm van achterstand kampen. Het grote aantal effectonderzoeken naar op jongeren gerichte interventies getuigt van het feit dat volgens de YEI-voorschriften dergelijke evaluaties op gezette tijden moeten worden uitgevoerd.

Ongeveer de helft van de CIE's die sinds 2015 in het kader van het evaluatiehulpdeskproject<sup>23</sup> zijn geïnventariseerd, heeft betrekking op thematische doelstelling 8 ("bevordering van duurzame en kwalitatief hoogstaande werkgelegenheid en ondersteuning van arbeidsmobiliteit"). Bij de meeste CIE's naar interventies die in het kader van TD 9 zijn gefinancierd, gaat de aandacht uit naar de vraag hoe effectief de interventies waren qua arbeidsmarktintegratie van kwetsbare werklozen (zoals bij de Duitse interventie die op langdurig werklozen was gericht en uit het operationele programma "ESF Bund Deutschland 2014-2020" werd gefinancierd).

De vormen van ondersteuning voor werklozen waarnaar het meeste onderzoek is gedaan, zijn opleidingen, stages en andere soorten werkervaring, en gesubsidieerde banen. Dit blijkt duidelijk uit de voorbeelden die in deze leidraad worden onderzocht (zie het voorbeeld van de Italiaanse evaluatie in Marche, de evaluatie van opleidingscursussen voor migranten in Duitsland, en de beroepsopleidingen voor personen die geen werk hebben en evenmin onderwijs of een opleiding volgen (NEET's) in Letland). Diensten op het gebied van loopbaanbegeleiding en arbeidsbemiddeling worden minder vaak onderzocht (zoals het Zweeds voorbeeld in deze leidraad, waarin een proefproject bij ODA's rond intensievere steun aan werklozen is geëvalueerd aan de hand van een gerandomiseerde benadering). Uit de gegevens van de helpdesk blijkt dat ESF-interventies voor de ondersteuning van zelfstandige arbeid of het oprichten van een bedrijf niet vaak worden geëvalueerd<sup>24</sup>.

In sommige gevallen worden met een CIE meerdere typen interventies tezamen onderzocht. Dat brengt het

<sup>23</sup> Zie [https://ec.europa.eu/regional\\_policy/en/policy/evaluations/member-states/](https://ec.europa.eu/regional_policy/en/policy/evaluations/member-states/)

<sup>24</sup> Enkele voorbeelden daarvan zijn: [Borik V. et al., 2015](#) – [Ires Piemonte, 2019](#) – [Openfield, 2019](#)

gevaar met zich mee dat verschillende interventielogica's vermengd raken en de betrouwbaarheid van de vergelijking tussen de behandelings- en de controlegroep afneemt.

Opvallend is een poging om de effecten van op kwetsbare personen gerichte interventies te evalueren. Daarbij ging het niet alleen om de resultaten op het gebied van werkgelegenheid, maar ook om soft outcomes: bij een CIE in Duitsland lag de nadruk op dat aspect en werd beoordeeld welk effect werkgelegenheidsprogramma's voor langdurig werklozen hadden op de gezondheidsbeleving, de tevredenheid met het leven, het gevoel van verbondenheid en indicatoren voor de maatschappelijke status.

ESF-interventies op onderwijsgebied worden minder vaak met een contrafeitelijke methode onderzocht. Het aantal CIE's dat te maken heeft met thematische doelstelling 10 is kleiner dan het aantal dat te maken heeft met thematische doelstelling 8 of 9, en de nadruk ligt vaak op interventies die worden gefinancierd in het kader van investeringsprioriteit 10.IV en die betrekking hebben op het beroepsonderwijs en de bijbehorende effecten op arbeidsmarktintegratie. Twee redenen hiervoor zijn de beperktere toegang tot datasets en de striktere privacyvoorschriften<sup>25</sup>. Toch zijn er wel pogingen in die richting ondernomen: in Spanje (Asturië) is een beoordeling uitgevoerd van een interventie op middelbare scholen om vroegtijdige schooluitval tegen te gaan; in Polen (Podalskie) werden met CIE's de effecten onderzocht van een project dat bedoeld was om jonge leerlingen te stimuleren om een beroepsopleiding te volgen; in Portugal werd onderzoek gedaan naar beurzen voor studenten in het hoger onderwijs.

CIE's kunnen op verschillende beleidsniveaus worden uitgevoerd (d.w.z. een of meer prioritaire assen, specifieke doelstellingen of concrete acties<sup>26</sup> van een programma) en kunnen betrekking hebben op nationale of regionale ESF+-programma's, gericht zijn op homogene doelgroepen (mannen of vrouwen, jongeren of kwetsbare bevolkingsgroepen, langdurig werklozen enz.) of op soorten interventies (bv. opleidingen, diensten in verband met sociale inclusie, of ontmoediging van voortijdige schooluitval (zie het vorige tekstkader)).

Voorbeelden uit de lidstaten wijzen erop dat tal van binnen het ESF+ gebruikte instrumenten geschikt zijn voor een CIE, waaronder opleidingen, werkgelegenheidsprikkels en arbeidsmarktdiensten (zoals loopbaanbegeleiding en coaching). Daar staat tegenover dat het lastiger is om een CIE uit te voeren naar interventies die te maken hebben met baanwisselingen en duobanen, met prikkels voor startende bedrijven of met steun voor systemen en structuren. Dit geldt ook voor interventies op het gebied van onderwijs en sociale inclusie.

Het is leerzaam om aan de hand van de volgende vragen na te gaan welke interventies in CIE-opzicht het meest veelbelovend zijn.

### **Is er sprake van een discrete, onderscheidende en relatief homogene interventie?**

*Duidelijk  
aanwijsbare  
behandeling*

De behandeling of behandelingen die bij de interventie worden uitgevoerd, moeten kunnen worden onderscheiden van de behandeling of behandelingen bij andere interventies. Bovendien moet er een zinvol onderscheid bestaan tussen de baten van de interventie voor de deelnemers en het nut van de interventie voor vergelijkbare groepen mensen. Een contrafeitelijke benadering is niet mogelijk of wenselijk wanneer een

<sup>25</sup> Zie bijvoorbeeld [Isméri Europa – Ecorys – Institute for Employment Studies, 2019](#)

<sup>26</sup> Artikel 2, punt 4, van Verordening (EU) 2021/1060 van het Europees Parlement en de Raad luidt: "concrete actie": a) een in het kader van de betrokken programma's gekozen project, overeenkomst, actie of groep projecten; b) bij financieringsinstrumenten, een bijdrage uit een programma aan een financieringsinstrument en de daaropvolgende financiële steun die door die financieringsinstrumenten aan eindontvangers wordt verleend.

behandeling zo slecht is afgebakend dat niet langer een afzonderlijke groep ontvangers kan worden aangewezen.

De CIE-methoden worden uiterst complex en ingewikkeld als de behandelstatus van een bepaalde eenheid (een onderneming of een individu) invloed heeft op het mogelijke resultaat van andere eenheden (ten gevolge van zogeheten bredere “algemene evenwichtseffecten”). Bij opleidingsprogramma’s kunnen deze effecten optreden als de personen die het programma hebben doorlopen het voor niet-cursisten moeilijk maken om op korte termijn werk te vinden. Wordt dit als een aanzienlijk probleem gezien (bijvoorbeeld bij grootschalige interventies), dan kan een macro-economische analyse nodig zijn om de omvang van de substitutie- en verplaatsingseffecten te onderzoeken. Wanneer dergelijke effecten zich dreigen voor te doen, moeten de beheerautoriteiten advies van deskundigen inwinnen.

De interventie zelf moet relatief homogeen zijn. Dit betekent dat alle deelnemers aan een interventie aan min of meer hetzelfde pakket maatregelen moeten worden onderworpen. Als het pakket maatregelen waarmee de deelnemers in het kader van één en dezelfde interventie worden geconfronteerd te zeer uiteenloopt, heeft dat een aantal consequenties voor het CIE. Ten eerste is het misschien niet logisch om van een samenhangende interventie te spreken omdat het feitelijk om interventies gaat waarbij afzonderlijke oorzakelijke processen spelen. Ten tweede is het moeilijk om een interpretatie te geven aan effecten die worden gerapporteerd als de gemiddelde netto-effecten van een groep uiteenlopende interventies. Ten derde kan er aanleiding zijn voor een subgroepanalyse, maar als de behandelingsgroep uit te veel subgroepen bestaat, kan het vanwege de beperkte steekproefomvang lastig zijn om bruikbare bevindingen te rapporteren.

*Homogene  
interventies*

### **Wordt de behandeling vergeleken met een situatie waarin geen behandeling plaatsvindt of bestaan er andere relevante behandelingen?**

Uit het ESF worden nationaal en regionaal arbeidsmarktbeleid en beleid op het gebied van sociale inclusie medegefinancierd. Bij CIE-gerelateerde evaluatieschema’s moet daarom zorgvuldig worden gelet op de vraag of de interventie duidelijk onderscheidbaar is en of mensen in het kader van andere (nationale of regionale) programma’s en financieringsbronnen diensten kunnen afnemen. Het is belangrijk dat de behandelingen die worden geëvalueerd daadwerkelijk verschil maken wat betreft de mogelijkheden of middelen die de deelnemers in vergelijking met de controle-eenheden ter beschikking staan en het is belangrijk dat die verschillen kunnen worden gemeten en gemonitord.

De problemen rond dergelijke “complexe behandelingen” zijn vaak contextafhankelijk en bemoeilijken het opzetten en uitvoeren van CIE’s. Die problemen onderstrepen hoe belangrijk het is dat de evaluatie vóór de uitvoering zorgvuldig wordt gepland (een evaluatieschema wordt opgezet).

*Complexe  
behandelingen*

### **Is er sprake van een grote en goed afgebakende doelgroep?**

Voor een CIE is, in vergelijking met andere evaluatievarianten, een grote steekproefomvang nodig. Het is cruciaal dat de doelgroepen uit voldoende aantallen mensen bestaan en er moeten controlegroepen van voldoende

*Grote  
steekproefomvang*

omvang kunnen worden gevonden. In het onderstaande wordt uitvoeriger op dit punt ingegaan.

Het is belangrijk dat een interventie die kandidaat is voor een CIE op een goed afgebakende groep is gericht. Het is moeilijk om een zinvolle controlegroep aan te wijzen als niet helder is uit welke mensen de doelgroep bestaat. Bij sommige interventies wordt bewust geprobeerd om mensen langs informele weg voor een behandeling te werven, waardoor processen worden gestimuleerd die niet vooraf zijn vastgesteld of overmatig prescriptief zijn (lastige maatschappelijke doelgroepen, zoals jonge NEET's of kansarme groepen, kunnen met behulp van incidentele of individuele procedures worden bereikt). Daardoor kan het moeilijk zijn om precies te bepalen wie de behandel eenheden zijn en wie tot de bijbehorende controlegroep behoort.

*De identiteit van de doelgroep vaststellen*

### **Bestaat er een duidelijk causaal mechanisme?**

Zoals eerder werd opgemerkt toen de belangrijkste evaluatiemethoden en de “interventiologica” werden geïntroduceerd, is het bij een theoriegebaseerde evaluatie vaak nuttig om deze voorafgaand aan of in combinatie met een CIE uit te voeren. Voor de opstellers van een CIE is het in meerdere opzichten nuttig om een uitgebreide interventiologica te formuleren, met name wanneer zij nagaan of aan de interventie een samenhangend causaal mechanisme ten grondslag ligt. In het geval van interventies zonder duidelijk en overtuigend causaal mechanisme is het twijfelachtig of er effecten van een zodanige omvang ontstaan dat die met een CIE statistisch kunnen worden vastgesteld.

*Onderscheidend beleidsmechanisme*

### **Kunnen de uitkomsten kwantitatief worden vastgesteld?**

Het is noodzakelijk over kwantificeerbare maatstaven voor de uitkomsten (of resultaten) te beschikken. Zulke gegevens en indicatoren kunnen afkomstig zijn uit administratieve bronnen of uit zeer gerichte enquêtes.

*Noodzaak van resultaatmeting*

In bepaalde situaties kunnen interventies beoogde resultaten opleveren die alleen kwantitatief kunnen worden gemeten als er specifieke voorzieningen worden getroffen. Een interventie kan bijvoorbeeld bedoeld zijn een verandering in attitudes, overtuigingen of meningen teweeg te brengen. In zo'n geval moet die verandering met behulp van een enquête worden gemeten. Voor sommige interventies geldt dat de resultaten tamelijk vaag of slecht zijn gedefinieerd. Nogmaals, een interventiologica kan helpen om dieper inzicht te verkrijgen in het beoogde doel van de interventie en in de wijze waarop met die interventie de verandering in de betreffende resultaten moet worden gerealiseerd.

### **Wordt de interventie zo geïntroduceerd dat een betekenisvolle controlegroep kan worden gevonden?**

Om een betekenisvolle controlegroep te kunnen aanwijzen, is het belangrijk om na te denken over de selectiemethode van behandel eenheden (personen of ondernemingen) voor een interventie of om na te gaan waarom zij tot deelname hebben besloten; na te gaan of voor de controlesteekproef en voor de behandelingsgroep dezelfde gegevensbron — bv. hetzelfde enquête-instrument — kan worden gebruikt; en, ten slotte, of het nodig is een controlesteekproef samen te stellen die met dezelfde arbeidsmarktsituatie te maken heeft als de behandelingsgroep. In het onderstaande tekstkader worden enkele voorbeelden nader belicht.

*Selectiemechanisme voor de behandeling*

Als een interventie verplicht is en min of meer gelijktijdig op de gehele doelpopulatie wordt toegepast, kan het moeilijk blijken om een niet-behandeld deel van de doelpopulatie te vinden dat als controlegroep kan fungeren.

### Tekstkader 5 Controlegroepen afbakenen

In de CIE-voorbeelden waarnaar in deze leidraad wordt gekeken, werd de keuze van de controlegroepen ingegeven door de interventiekenmerken (zoals selectiecriteria) en door de beschikbaarheid van geschikte gegevens.

In vergelijking met de CIE's uit de vorige programmeringsperiode gebeurde de afbakening en selectie van de controlegroep vaker aan de hand van administratieve gegevens, en met name van werklozenregisters. De strategie die het meest wordt gehanteerd is dat wordt gezocht naar kandidaten voor de controlegroep met vergelijkbare kenmerken die, conform de selectiecriteria, gedurende een bepaalde periode bij een ODA als werkloze staan ingeschreven. Dit geldt voor een aantal evaluaties in Italië (Marche, de provincie Trento en Piëmont), Polen (Lubelskie en Podlaksie), evenals voor de Duitse evaluaties van opleidingscursussen voor migranten, werkgelegenheidsprogramma's en maatregelen voor de arbeidsmarktintegratie van werklozen in Baden-Württemberg.

Bij andere evaluaties, die weliswaar niet aan bod komen in de voorbeelden die uitvoerig in deze leidraad worden beschreven, werden mensen die zich wel hadden aangemeld maar niet voor de behandeling waren geselecteerd naderhand in de controlegroep ondergebracht. Dit gebeurde bij de Italiaanse evaluatie van de jongerengarantie, de Duitse evaluatie van het in de periode 2007-2013 gefinancierde proefproject "Bürgerarbeit" en de evaluatie van het Pipol-programma dat in Friuli Venezia Giulia werd uitgevoerd<sup>27</sup>. Deze strategie wordt echter niet vaak toegepast, want de monitoringinformatiesystemen van de meeste landen en regio's bevatten geen informatie over mensen die zich wel hadden aangemeld maar niet hebben deelgenomen.

De Duitse evaluatie van op langdurig werklozen gerichte interventies<sup>28</sup> kende een andere aanpak, want in die evaluatie werd de *intention to treat* (ITT) beoordeeld en niet de *average treatment effects on treated* (ATT). De behandelingsgroep bestond in dit geval uit mensen die gedurende de gehele uitvoeringsperiode van het programma mogelijk aan de criteria voldeden, ongeacht of zij al dan niet daadwerkelijk werden behandeld (vanaf 2015), en de controlegroep bestond uit mensen die aan diezelfde selectiecriteria voldeden, maar die vóór de uitvoering van het programma (2010-2012) langdurig werkloos waren geweest.

Alleen in het Zweedse voorbeeld, waarbij een gerandomiseerde aanpak werd gehanteerd, werd de controlegroep op aselechte wijze samengesteld: de behandeling (intensievere ondersteuning door ODA's) werd met opzet op aselechte wijze aangeboden aan een groep jongeren, terwijl de controlegroep van de ODA's de reguliere ondersteuning kregen.

In het handjevol evaluaties van ESF-maatregelen voor ondernemingen werden de behandelings- en controlegroep afgebakend door onderscheid te maken tussen kandidaten met en zonder financiering, net als bij sommige Deense evaluaties waarin de prestaties van uit het ESF gefinancierde ondernemingen werden vergeleken met de prestaties van een steekproef van ondernemingen die qua kenmerken wel vergelijkbaar waren maar geen steun hadden ontvangen<sup>29</sup>.

<sup>27</sup> Zie respectievelijk [Isfol, 2016](#) – [IAW Institut für Angewandte Wirtschaftsforschung – ISG Institut für Sozialforschung und Gesellschaftspolitik, GmbH, 2015](#) – [Ismeri Europa, 2018](#)

<sup>28</sup> Zie [Boockmann B. et al., 2019](#)

<sup>29</sup> Bijvoorbeeld: [Danmarks Statistik et al., 2017](#) en [Danmarks Statistik et al., 2018](#)

## 2.2. De evaluatievragen en uitkomstvariabelen



### 2.2.1. Wat zijn de doelstellingen van de interventie?

Bij het opstellen van een evaluatieschema verdient het voor alles aanbeveling om de doelstellingen en belangrijkste kenmerken van de interventie te beschrijven.

*Het CIE-model combineren met inzichten uit de interventielogica*

In veel gevallen bestaan er al stukken met een beschrijving van de doelstellingen van de interventie. Bij CIE's is het echter belangrijk om te preciseren welke resultaten en veranderingen met de interventie worden beoogd en dus welke effecten worden verwacht.

Doorgaans is het nuttig om een interventielogica te formuleren waarin staat met welke middelen de diverse inputs en activiteiten van de interventie moeten worden gekoppeld aan de outputs, uitkomsten (of resultaten) en dus aan de effecten.

### 2.2.2. Waartoe dient de evaluatie?

Bij het opzetten van een evaluatieschema voor een CIE is het belangrijk goed na te denken over het doel van de evaluatie. De kans is klein dat de evaluatie het benodigde feitenmateriaal oplevert als er geen duidelijk beeld bestaat van de noodzaak van de evaluatie. Bij evaluaties van uit het ESF gefinancierde interventies moet worden stilgestaan bij een reeks vragen:

- Wat is het doel en het karakter van de evaluatie in het licht van de regelgevingseisen en richtsnoeren van de Europese Commissie?
- Wie zijn de belangrijkste belanghebbenden bij de evaluatie?
- Wat gebeurt er met de evaluatieresultaten?
- Op welke concrete vragen moet de evaluatie een antwoord bieden?

### Wat zijn de doelstellingen en het karakter van de evaluatie?

Als eerste moet worden **gemotiveerd** waarom het nodig is om een evaluatie uit te voeren. In Verordening (EU) 2021/1060 is bepaald: "De lidstaat of de beheerautoriteit verricht evaluaties van de programma's met betrekking tot een of meer van de volgende criteria: effectiviteit, efficiëntie, relevantie, samenhang en Uniemeerwaarde, dit ter verbetering van de kwaliteit van het ontwerp en de uitvoering van de programma's. De evaluaties kunnen ook betrekking hebben op andere relevante criteria, zoals inclusiviteit, non-discriminatie en zichtbaarheid, en op meer dan één programma"<sup>30</sup>. Zoals hierboven is gebleken, hebben de bevindingen van een CIE meestal betrekking op de effectiviteit (in hoeverre zijn de verwachte uitkomsten

*De doelstellingen en het karakter van de evaluatie*

<sup>30</sup> Artikel 44, lid 1, van Verordening (EU) 2021/1060.

behaald) en de efficiëntie (de kosteneffectiviteit of de kosten-batenverhouding van de interventie).

Meer in het algemeen moedigt de Commissie de lidstaten aan om bij de beleidsvorming resultaatgerichte benaderingen te volgen en evaluaties uit te voeren die qua toepassingsgebied, opzet en tijdschema voldoen aan de behoeften van de lidstaten zelf. Binnen de ESF+-programma's kunnen wat dat betreft ook CIE's worden uitgevoerd om een antwoord op specifieke evaluatievragen te krijgen of om te voldoen aan het nationale evaluatiebeleid.

Als tweede moet het **karakter van de evaluatie** worden bepaald:<sup>31</sup>

- Met **effectgeoriënteerde** evaluaties worden de effecten van een programma of groep programma's onderzocht in het licht van nationale en EU-prioriteiten (daarbij kan het gaan om de macro-economische effecten van het ESF, om bepaalde beleidslijnen en thema's die centraal staan of om horizontale prioriteiten, zoals kinderen en gelijke kansen).
- Met een **proces- of uitvoeringsgeoriënteerde** evaluatie wordt de uitvoering van een programma ondersteund, worden de vooruitgang en de uitvoeringsmethoden geanalyseerd en aanbevelingen gedaan voor verbeteringen van het programma.

In beginsel geldt dat voor effectonderzoeken de contrafeitelijke methode kan worden gehanteerd, terwijl voor procesevaluaties andere methoden nodig zijn (zie tevens figuur 1 hierboven).

In tegenstelling tot de programmeringsperiode 2014-2020 schrijft Verordening (EU) 2021/1060 houdende gemeenschappelijke bepalingen (hierna de "GB-verordening" genoemd) voor de programmeringsperiode 2021-2027 geen specifiek aantal effectonderzoeken voor. In plaats daarvan wordt in de VGB een evaluatiestrategie verlangd waarmee kan worden nagegaan op welke wijze de steun uit de Europese fondsen heeft bijgedragen tot de verwezenlijking van de doelstellingen van het programma met betrekking tot alle belangrijke strategische profielen<sup>32</sup>. In de GB-verordening is evenmin bepaald op welke prioriteiten of interventies de evaluaties gericht moeten zijn en wordt het besluit daarover aan de beheerautoriteit en het evaluatieplan gelaten. Dit betekent dat in de evaluatiestrategie van elk afzonderlijk programma de mix van effect- en procesevaluaties moet worden beschreven en moet worden verduidelijkt op welke prioriteiten en interventies die strategie in hoofdzaak is gericht. Uiterlijk in juni 2029<sup>33</sup> moet er een definitief en generiek effectonderzoek zijn verricht, maar voor de overige evaluaties die in het evaluatieplan beschreven moeten worden, gelden geen andere tijdsbeperkingen.

### Voor wie is de evaluatie in hoofdzaak bestemd?

Vastgesteld moet worden voor wie de evaluatie is bestemd. Naargelang het karakter van de evaluatie kan het gaan om beleidsmakers, beheerautoriteiten en programmabeheerders, andere beheerautoriteiten of uitvoeringsorganen in de lidstaat, of om nationale of regionale autoriteiten die soortgelijke programma's beheren. Wanneer er gegevens worden aangeleverd door instellingen die niet onder het programmabeheer vallen, dan moeten die

*Belanghebbenden  
in kaart brengen*

<sup>31</sup> [Zie Europese Commissie, 2007](#)

<sup>32</sup> Artikel 44, lid 1, GB-verordening.

<sup>33</sup> Artikel 44, lid 2, GB-verordening.



organen (“gegevenseigenaren”) ook als belanghebbenden worden aangemerkt. Om te komen tot gezamenlijke verantwoordelijkheid voor het opzetten en uitvoeren van de evaluatie is het belangrijk dat alle belangrijke belanghebbenden zitting hebben in een evaluatiestuurgroep. Dat geldt ook voor een aantal evaluatiedeskundigen uit de academische wereld of van openbare instellingen die de beheerautoriteit van technische adviezen voorzien.

### **Wat gebeurt er met de evaluatieresultaten?**

Wanneer is vastgesteld voor wie de evaluatie bestemd is, kan worden bepaald hoe de bevindingen zullen worden gebruikt. Hieraan kan handen en voeten worden gegeven door de stuurgroep te betrekken bij het formuleren van de evaluatievragen, en door besprekingen te houden over de inhoud van de opdracht.

De resultaten van een CIE spelen vaak een rol bij besluiten over twee belangrijke vragen:

- of een bestaande interventie moet worden voortgezet; en
- of een nieuw type interventie breder moet worden toegepast (dus in opgeschaalde vorm).

In het geval van de eerste vraag kan, als er sprake is van druk op de begroting, worden geprobeerd met een CIE de effectiviteit van een bestaand of lopend programma te beoordelen; aan de betrokken middelen kan eventueel een andere bestemming worden gegeven. In die situatie ligt het niet voor de hand dat de interventie al eerder met een contrafeitelijk effectonderzoek is geëvalueerd.

In het geval van de tweede vraag spelen er bij de interventies mogelijk beperkingen op uitvoerend vlak: de interventie is bijvoorbeeld in een bepaalde regio of in een bepaald gebied van een lidstaat uitgevoerd, of slechts voor een beperkte periode. In dergelijke situaties kan aan de hand van de resultaten van een CIE worden vastgesteld of de betreffende interventie effectief is en of het nut heeft deze elders uit te voeren. In zulke gevallen gaat het om proefinterventies die later breder worden uitgerold.

### **Op welke vragen moet een antwoord komen?**

Als de doelstellingen van de interventie en het doel en de uiteindelijke toepassingen van de evaluatie eenmaal zijn vastgesteld, en een duidelijke doelgroep is aangewezen, is het mogelijk gedetailleerder te bepalen op welke vragen het CIE moet ingaan. Vaak zijn er uiteenlopende doelgroepen en belanghebbenden met vragen van causale aard die zij met een CIE onderzocht willen zien. De beheerautoriteit of de beoordelaar moet deze vragen door middel van een uitgebreide raadpleging inventariseren en rekening houden met de diverse standpunten en suggesties van de mensen die bij de verschillende interventiestadia betrokken zijn. Vervolgens moet de beheerautoriteit en/of de beoordelaar de prioriteit van de vragen bepalen en het CIE op de belangrijkste en meest geschikte vragen richten.

Hieronder volgen enkele aspecten die bij het opstellen van een definitieve lijst met centrale onderzoeksvragen voor een CIE in overweging kunnen worden genomen:

- Heeft de interventie op de korte, middellange en lange termijn de beoogde resultaten opgeleverd of bijgedragen tot die resultaten? Plus: verschillen

*Belangrijkste  
onderzoeksvragen*

de kortetermijneffecten significant van de langetermijneffecten? Er moet prioriteit worden gegeven aan vragen die op deze aspecten ingaan.

- Kunnen de uitkomsten kwantitatief worden gemeten? Aan welke bron ook de voorkeur wordt gegeven, het CIE moet gebaseerd zijn op een geschikte gegevensreeks (administratieve gegevens of een directe enquête die naar de deelnemers is verstuurd).
- In hoeverre zijn de veranderingen in de omstandigheden van de deelnemers of in de sociaaleconomische context toe te schrijven aan de interventies? Bij dit soort vragen is het noodzakelijk om de omvang van de netto-effecten van de interventie te meten, zodat een vergelijking kan worden gemaakt met de deelnemers aan de controlegroep en met de contextindicatoren.
- Ging het bij alle leden van de doelgroep om dezelfde interventie-effecten? Had een op langdurig werklozen gerichte interventie bijvoorbeeld hetzelfde effect op mannen als op vrouwen? En op mensen jonger dan 25 of ouder dan 50 jaar? De effecten op verschillende subgroepen laten zich met een CIE uitstekend onderzoeken wanneer het aantal personen in de behandelings- en de controlegroep groot genoeg is.
- Was de interventie kosteneffectief (vergeleken met alternatieven)? Wat is de kosten-batenverhouding van de interventie? Voor deze efficiëntieanalyse moeten de effecten met een CIE nauwkeurig kunnen worden gemeten en moet die CIE exacte informatie opleveren over de directe en indirecte kosten en baten.
- Is er ruimschoots voldoende informatie en kennis over de effecten van soortgelijke interventies? In hoeverre kan deze kennis worden toegepast op de onderzochte interventie? Als het om bekend beleid gaat waarvan de effecten uitgebreid zijn onderzocht, kan het nuttig zijn om de evaluatievragen op specifieke beleidsaspecten te richten om te voorkomen dat andere analyses worden overgedaan. Een uitvoerige bestudering van de bestaande literatuur kan aanknopingspunten opleveren voor de besluiten hieromtrent en biedt in ieder geval nuttige houvast bij het opzetten van het CIE.

Voordat opdracht tot een CIE wordt verleend, is het belangrijk om een duidelijk beeld te hebben van de diverse onderzoeksvragen waarop het CIE antwoord moeten bieden. Voor elk evaluatieschema geldt dat het cruciaal is om te bespreken welke vragen in de evaluatie zullen worden gesteld.

Het is belangrijk om de prioriteit van de vragen te bepalen en weerstand te bieden aan de heersende trend om evaluaties te overladen met veel te veel vragen. Het is een hele puzzel om het juiste midden te vinden tussen enerzijds een evaluatie die relevant is voor belanghebbenden met uiteenlopende belangen en anderzijds een evaluatie die hanteerbaar is. Een evaluatie kan stuurlaos worden als de onderzoeksvragen te zeer uiteenlopen, wat uitmondt in een evaluatie waarin tal van punten op een suboptimale manier worden aangesneden. Omdat vaak geldt “minder is meer”, is de prioriteringsfase van het planningsproces van een evaluatie cruciaal.

*De prioriteit van vragen bepalen*

Voor het bepalen van de prioriteit van de evaluatievragen moeten dublures worden geschrapt en dient op basis van relevante beginselen aan elke vraag een score te worden toegekend. Daarbij kan het gaan om: het belang en de feitelijke inzet van de belanghebbende die de vraag heeft geformuleerd, de mate waarin de vraag aansluit bij de interventielogica van het programma, de

relevantie van de vraag gelet op het algemene doel van de evaluatie en de haalbaarheid van de vraag in het licht van de beschikbare gegevens, tijd en middelen<sup>34</sup>. Dit resulteert in een lijst waarop de vragen in volgorde van belang staan gerangschikt en de belangrijkste vragen kunnen worden geselecteerd.

Soms is het mogelijk een reeks evaluatievragen onder te brengen in een meer algemene vraag, zoals in de vraag: “wat was het netto-effect van de interventie?” Daarin kunnen andere vragen worden ondergebracht, bijvoorbeeld over het effect op verschillende groepen deelnemers of eventueel zelfs over het effect op verschillende momenten. Dit moet echter te allen tijde een behapbare en bruikbare reeks vragen opleveren.

### ***Welke evaluatiecriteria kunnen er aan de evaluatie worden gekoppeld?***

Het verband tussen de evaluatiecriteria en de evaluatievragen is hierboven al aangestipt, maar verdient nadere toelichting. Evaluatiecriteria (efficiëntie, effectiviteit, EU-meerwaarde, samenhang enz.) zijn nodig om een waarde te kunnen toekennen aan het verzamelde feitenmateriaal en tot een beoordeling van het beleid te kunnen komen (efficiëntie enz.). De evaluatievragen zijn op hun beurt nodig om de behoeften van de opdrachtverlener te expliciteren en de aandacht te vestigen op de belangrijkste beleidskwesties waar het om gaat. Toch hangen de evaluatiecriteria en de evaluatievragen met elkaar samen. In de regel is elke vraag terug te voeren op een specifiek criterium. Dit verband is voor zowel de beheerautoriteiten als voor de beoordelaars belangrijk omdat daarmee de evaluatieopzet — die noodzakelijkerwijs op de vragen berust — wordt gekoppeld aan de vereisten van de GB-verordening 2021-2027, die op een andere manier verbonden zijn met de evaluatiecriteria.

*Evaluatiecriteria en evaluatievragen*

Hieronder staan enkele voorbeelden van veelvoorkomende evaluatievragen, vergezeld van de bijpassende criteria. Het gaat om een keuze uit en bewerking van de evaluatievragen uit de “Toolbox voor betere regelgeving”<sup>35</sup> (instrument nr. 47 inzake evaluatiecriteria en -vragen) van de Europese Commissie, met als bedoeling voorbeelden te geven van vragen waarvoor een effectanalyse nodig is en zo mogelijk een CIE:

- *Voorbeelden van veelgestelde vragen over de effectiviteit*
  - Wat waren de kwantitatieve effecten van de interventie?
  - In hoeverre zijn die veranderingen/effecten te danken aan de interventie?
  - In hoeverre kunnen factoren die van invloed zijn op de waargenomen verwezenlijkingen in verband worden gebracht met de EU-interventie?
- *Voorbeelden van veelgebruikte vragen over de efficiëntie*
  - In hoeverre was de interventie kosteneffectief?
  - In hoeverre waren de kosten van de interventie gerechtvaardigd, gelet op de daarmee gerealiseerde veranderingen/effecten?
  - Als er tussen gebieden aanzienlijke verschillen in kosten (of baten) bestaan, wat is daar dan de oorzaak van? Wat is het verband tussen die verschillen en de interventie?
- *Veelvoorkomende vragen over de EU-meerwaarde*
  - Wat is de meerwaarde van de ESF+-interventie(s) afgezet tegen de resultaten van soortgelijke nationale en/of regionale interventies?

<sup>34</sup> Zie bijvoorbeeld [Centre for Disease Control and Prevention CDC, 2013](#)

<sup>35</sup> Zie [Europese Commissie, 2017](#)

- Welke gevolgen zou het stopzetten of intrekken van de bestaande ESF+-interventie naar alle waarschijnlijkheid hebben?

## 2.3. Gegevens waarmee de controlegroep kan worden bepaald en de uitkomstvariabelen kunnen worden gemeten



### 2.3.1. Zijn de juiste gegevens beschikbaar of kunnen die beschikbaar komen?

Uit besprekingen met beheerautoriteiten en evaluatiedeskundigen uit de gehele EU blijkt dat toegang tot de juiste gegevens een van de belangrijkste uitdagingen is bij het uitvoeren van CIE's. Een belangrijke praktische vraag om bij stil te staan is of de vereiste gegevens wel beschikbaar zijn. In dit gedeelte staat een vereenvoudigde categorisering van de benodigde soorten gegevens, plus een bespreking van de bronnen waaruit dergelijke gegevens kunnen worden verkregen en van de stappen die met het oog op de verzameling van primaire gegevens eventueel moeten worden gezet. Ook wordt ingegaan op het essentiële thema gegevensbescherming.

Met betrekking tot goede planning moet een belangrijk punt worden opgemerkt. Vroeger kwam het wel voor dat de uitvoering van CIE's stukliep op een gebrek aan gegevens, omdat was verzuimd in een vroegtijdig stadium adequate plannen te ontwikkelen. Het is bij bestaande interventies belangrijk om te bepalen wie er tot de groep met en de groep zonder behandeling behoren en om mechanismen op te zetten waarmee de gegevens over die personen worden verzameld. Zij staan immers centraal in de evaluatie. Om ervoor te zorgen dat op de juiste momenten de juiste gegevens worden verzameld, moeten bij nieuwe interventies in een vroeg stadium stappen worden ondernomen.

*Gegevensverzameling plannen*

#### Welke soorten gegevens zijn nodig?

Voor een CIE zijn over het algemeen drie soorten gegevens nodig. In sommige gevallen kan een bron één of meer van die soorten gegevens bevatten. Het betreft: gegevens over de behandelings- en controlegroep, uitkomstgegevens en contextuele gegevens.

*Belangrijkste gegevensbronnen*

- **Gegevens over de behandelings- en controlegroep:** er zijn gegevensbronnen nodig waarmee de beoordelaars de afzonderlijke eenheden van behandelings- en controlegroep kunnen bepalen (ondernemingen, personen of eventueel geografische gebieden).
- **Uitkomstgegevens:** zoals de figuren 6 en 7 in hoofdstuk 3 van deze leidraad laten zien, moeten bij een CIE de uitkomsten voor zowel de behandelings- als de controlegroep kunnen worden gemeten. Idealiter worden de uitkomstgegevens voor beide groepen met behulp van

dezelfde methode verzameld en worden de resultaten op dezelfde momenten gemeten.

- **Contextuele gegevens:** er zijn gegevens nodig om een goed gelijkende controle- en behandelingsgroep samen te stellen en om in de analyse voor eventuele resterende verschillen tussen die twee groepen te kunnen controleren. Het is belangrijk dat zoveel mogelijk gegevens worden verzameld over de kenmerken van de eenheden en over de factoren die mogelijk verband houden met de keuze om deel te nemen aan een interventie en met de potentiële resultaten. Dit geldt met name voor resultaatindicatoren die vóór de behandeling worden gemeten. Bij contextuele gegevens kan het ook gaan om gegevens die een beschrijving geven van lokale arbeidsmarkten (zoals lokale werkloosheidspercentages of krapte op de arbeidsmarkt) en gegevens waarmee een analyse per subgroep kan worden uitgevoerd.

In tabel 2 hieronder staan die drie soorten gegeven vermeld, met suggesties voor bronnen waar die gegevens eventueel kunnen worden opgevraagd. In het eerste tekstkader hieronder staan voorbeelden van gegevens die voor CIE's in het kader van het ESF worden gebruikt. Het tweede tekstkader bevat voorbeelden van een geïntegreerde dataset die wordt gebruikt bij het analyseren van arbeidsmarktfragmenten en voor evaluaties van het arbeidsmarktbeleid.

**Tabel 2 Soorten gegevens en gegevensbronnen**

Soorten gegevens	Bronnen
<b>Gegevens over de behandelingsgroep</b>	<ul style="list-style-type: none"> <li>- gegevens over de deelname aan interventies (veelal in bezit van begunstigden);</li> <li>- ESF+-monitoringgegevens (interventiekenmerken, start- en einddata, achtergrondgegevens, aanmeldgegevens).</li> </ul>
<b>Gegevens over de controlegroep</b>	<ul style="list-style-type: none"> <li>- administratieve gegevens over bijvoorbeeld sociale zekerheid, onderwijs en werkloosheidsuitkeringen;</li> <li>- gegevens over deelname (personen die wel voor deelname in aanmerking kwamen maar om andere redenen toch niet hebben deelgenomen<sup>36</sup>);</li> <li>- bestaande nationale enquêtes, zoals de arbeidskrachtenenquête.</li> </ul>
<b>Resultaatgegevens (benodigd voor zowel de behandelings- als de controlegroep)</b>	<ul style="list-style-type: none"> <li>- administratieve gegevens: gegevens over sociale zekerheid en werkloosheid kunnen bv. ook worden gebruikt om uitkomstmaten op te stellen (sociale en andere uitkeringen), gegevens over volksverzekeringen en fiscale gegevens (inkomsten en arbeidssituatie);</li> <li>- administratieve onderwijsgegevens (gestandaardiseerde toetsen van vaardigheden, slagingspercentages, inschrijving en aanwezigheid);</li> <li>- officiële bedrijfsregisters of fiscale administratie (productiviteit of omzet voor en na interne opleidingen of indiensttreding van nieuw personeel);</li> </ul>

<sup>36</sup> Het betreft deelnemers die wel in aanmerking kwamen maar niet zijn toegelaten; niet-toelating heeft vaak met stopzetting van de financiering te maken of met andere externe oorzaken (ziekte van lesgevers, vervoersproblemen enz.) Omdat de betrokkenen aan dezelfde criteria voldeden als de daadwerkelijke deelnemers en even gemotiveerd waren, kunnen zij een voorkeurscontrolegroep vormen; soms is hun aantal echter te gering om een controlegroep te vormen of zijn hun gegevens niet geregistreerd.

Soorten gegevens	Bronnen
	<ul style="list-style-type: none"> <li>- werkgelegenheids- of productiegegevens uit officiële statistische bron (om bij gebiedsgebonden contrafeitelijke analyses de omvang van de werkgelegenheid en het bbp te kunnen meten);</li> <li>- speciale enquêtes onder de behandelings- en controlegroep.</li> </ul>
<p><b>Contextuele gegevens/controlevariabelen (benodigd voor zowel de behandelings- als de controlegroep)</b></p>	<ul style="list-style-type: none"> <li>- administratieve stelsels, bv. uitkeringsgegevens die inzicht geven in de uitkeringsgeschiedenis vóór de behandeling; gegevens over volksverzekeringen en fiscale gegevens, gegevens over inkomens- en arbeidsverleden;</li> <li>- officiële arbeidsmarkt- of onderwijsstatistieken (bv. de arbeidskrachtenenquête, basisgegevens op regionaal of nationaal niveau die ook microgegevens op individueel niveau bevatten voor specifieke verfijning);</li> <li>- enquêtes onder de behandelings- en controlegroep. Wanneer de behandelregels duidelijk zijn, kunnen de controlegroepen van tevoren worden vastgesteld en is het mogelijk referentiegegevens te verzamelen;</li> <li>- monitoringsystemen — in sommige omstandigheden kunnen met monitoringsystemen meetgegevens over zowel de behandelings- als de controlegroep worden verzameld, bijvoorbeeld in het geval van aanmeldsystemen waarbij afgewezen kandidaten als controle-eenheden worden ingezet.</li> </ul>

### Tekstkader 6 Voorbeelden van gegevens die voor CIE's worden gebruikt

Van de voorbeelden die in deze leidraad uitvoerig worden bestudeerd, geldt waarschijnlijk voor de Duitse evaluaties dat daarvoor de meest uitvoerige administratieve datasets zijn gebruikt. In de geïntegreerde beroepsbiografieën (*Integrierten Erwerbsbiografien*, IEB) is een aantal administratieve informatiebronnen geïntegreerd die door de federale arbeidsdienst (*Bundesagentur für Arbeit*) worden beheerd en informatie bevatten over de loopbaan van mensen (zelfstandigen uitgezonderd), over perioden van werkloosheid, over sociaal-demografische kenmerken, over overdrachten (werkloosheidsuitkeringen) en over deelname aan actief arbeidsmarktbeleid. De IEB's kennen een lange geschiedenis; de ontwikkeling en het beheer van de dataset vergt aanhoudende inspanningen en investeringen. In tekstkader 7<sup>37</sup> staat meer informatie over de IEB's. In het geval van de Duitse evaluatie van werkgelegenheidsprogramma's werden deze gegevens aangevuld met een enquête in drie rondes met als bedoeling tijdens de verschillende stadia van de interventies informatie te verzamelen over de "soft outcomes" (naar de beleving van de deelnemers).

Voor de Italiaanse evaluaties die in deze leidraad uitvoerig worden onderzocht, zijn vergelijkbare gegevens gebruikt. Voor het samenstellen van mogelijke controlegroepen werden archieven met bij ODA's ingeschreven werklozen gebruikt. Met behulp van administratieve gegevens over veranderingen op de arbeidsmarkt (*Comunicazioni obbligatorie* — COB), waarin alle arbeidsovereenkomsten van ondernemingen en openbare werkgevers met uitzondering van zelfstandigen worden geregistreerd, werden de uitkomstvariabelen gemeten. Omdat de COB's op regionaal niveau worden beheerd, is de kwaliteit van de informatie wisselend. De laatste jaren zijn echter landelijke normen ingevoerd en is de kwaliteit verbeterd. Bij één evaluatie (voor de provincie Trento) kon de beoordelaar deze gegevens combineren met door het INPS (nationaal instituut voor de sociale zekerheid) verstrekte gegevens uit belastingaangiften en zodoende de inkomenseffecten op afzonderlijke personen meten.

In Polen werden controlegroepen samengesteld met behulp van administratieve gegevens uit de werklozenregisters van ODA's. Deze gegevens bevatten echter geen informatie over iemands arbeidsstatus of arbeidsverleden. Voor de evaluaties moest worden uitgeweken naar een alternatieve maatstaf (intrekking van de status van werkloze bij de ODA's) of moest door middel van enquêtes informatie

<sup>37</sup> Een vergelijkbaar voorbeeld is de Ierse databank "Jobseekers Longitudinal Dataset" (JLD), waarmee in één databank informatie wordt geïntegreerd over sociale uitkeringen en sociale verzekeringen, arbeidsmarktprogramma's en arbeidsverleden. De JLD werd onder meer gebruikt voor de evaluatie van het in de programmeringsperiode 2007-2013 gefinancierde activeringsprogramma JobBridge. Zie [Indecon, 2016](#)

worden verzameld onder een steekproef van zowel behandelde als niet-behandelde personen.

In Letland slaagde de beoordelaar erin twee belangrijke administratieve datasets samen te voegen. De gegevens van het Letse nationale arbeidsbureau (SEA) verschaften informatie over deelnemers en niet-deelnemers die op bepaalde momenten als werkloos stonden geregistreerd. De gegevens van de belastingdienst boden informatie over de arbeidssituatie en het inkomen van personen op verschillende momenten. Hierdoor kon de beoordelaar zowel de inkomenseffecten onderzoeken als het effect op de kans om op bepaalde momenten een baan te hebben.

In het enige geval van randomisatie (het voorbeeld uit Zweden) werd gebruikgemaakt van administratieve gegevens om de uitkomstvariabelen te meten en van een enquête onder arbeidsbureaus en arbeidsbemiddelingsbureaus om de intensiteit en het soort verleende steun in kaart te brengen.

Bij de evaluaties op onderwijsgebied werd gebruikgemaakt van administratieve gegevens om de controlegroep te bepalen en de uitkomstvariabelen te beoordelen. Dit geldt niet voor Polen, dat gebruik moest maken van een enquête om de variabelen te meten die als uitkomsten werden gebruikt.

Alles bij elkaar blijkt uit de voorbeelden hoe belangrijk juiste administratieve gegevens zijn voor CIE's. Idealiter worden zij zowel voor het bepalen van controlegroepen als voor het meten van de uitkomstvariabelen gebruikt. Gelet daarop moeten beheerautoriteiten die een CIE plannen er van tevoren op toezien dat de voor het uitvoeren van evaluaties benodigde administratieve gegevens beschikbaar zijn en moeten zij de juiste maatregelen nemen om eventuele problemen rond toegang tot en integratie van gegevens of andere hoofdbreken te verhelpen.

## Tekstkader 7 Voorbeelden van geïntegreerde databanken voor CIE's

### De *Jobseekers Longitudinal Dataset (JLD)* in Ierland

De *Jobseekers Longitudinal Dataset (JLD)* is een administratieve databank die wordt beheerd door het ministerie van Sociale Bescherming (*Department of Social Protection, DSP*).

De JLD is een ambitieuze poging om administratieve gegevens geschikt te maken voor onderzoeksdoeleinden. De ontwikkeling van de JLD begon ongeveer tien jaar geleden, nadat het DSP het *University College Dublin* opdracht had gegeven om een voorstudie uit te voeren naar het beheer van het *Live Register*<sup>38</sup> en naar arbeidsmarktgegevens in bredere zin. In de studie werden verschillende voorstellen gedaan om de gegevensverzameling te verbeteren en werd een aantal problemen geconstateerd (zoals gegevens die in meerdere gegevenssystemen voorkwamen, ontbrekende informatie enz.).

In de JLD zijn meerdere informatiebronnen geïntegreerd: betalingsgegevens en administratieve gegevens van het DSP, zoals uitkeringen in het kader van sociale bijstand en sociale zekerheid voor de beroepsbevolking die in het *Live Register* zijn opgenomen en gegevens met betrekking tot door het DSP beheerde beleidsprogramma's voor arbeidsmarktactivering; gegevens over arbeidsmarktprogramma's die worden beheerd door Solas (het nationale orgaan voor onderwijs en opleiding); gegevens die door de belastingdienst (*Revenue Commissioners*) zijn verzameld. De interventies die uit het ESF worden gefinancierd, vallen ten dele onder de JLD; meer in het bijzonder gaat het om de door het DSP en Solas gefinancierde interventies<sup>39</sup>.

De gegevens uit de bovengenoemde bronnen worden ingedeeld als een reeks tijdvakken. Een tijdvak begint wanneer iemand werkloos wordt en eindigt wanneer die persoon een baan vindt, een andere vorm van bedrijvigheid begint of deelneemt aan een opleidingsprogramma. Wanneer de status van die persoon opnieuw verandert, breekt een nieuw tijdvak aan. Het begin van een periode van werkloosheid valt samen met het moment waarop iemand een *Jobseeker's Benefit* of een *Jobseeker's Allowance* aanvraagt<sup>40</sup>. Een

<sup>38</sup> Het *Live Register* bevat informatie over mensen die zich op een lokaal kantoor van het DSP melden voor een *Jobseeker's Benefit (JB)* of *Jobseeker's Allowance (JA)* of aanspraak maken op verschillende andere wettelijke rechten.

<sup>39</sup> Programma's op het gebied van justitie (de *Youth Diversion Projects*, de *Young Persons Probation Projects* en *Integration and Employment of Migrants*), onderwijs (*Third Level Access* en *Adult Literacy*), wijkontwikkeling (de *Community Training Centres*) en defensie (het *Defence Forces Employment Support Scheme*) zijn niet opgenomen in de JLD.

<sup>40</sup> De *Jobseeker's Benefit* is een wekelijkse uitkering van het DSP voor mensen die zonder werk zitten, volledig werkloos zijn of in deeltijd werken omdat hun werkgever de arbeidsduur heeft verkort. Voor de JB komen werkzoekenden in aanmerking die voldoende sociale premies (PRSI) hebben afgedragen, die in het betreffende belastingjaar voldoende premies hebben betaald en die in totaal ten minste 104 keer premie hebben afgedragen. De *Jobseeker's Allowance* is een inkomensafhankelijke uitkering voor werkzoekenden die werkloos zijn en niet in aanmerking komen voor de *Jobseeker's Benefit* of daar niet langer aanspraak op

voordeel van deze indeling in tijdvakken is dat in de JLD aaneengesloten perioden van Jobseeker's Benefit (JB) en Jobseeker's Allowance (JA) met elkaar kunnen worden gekoppeld en als één tijdvak van werkloosheid kunnen worden weergegeven. Opgemerkt moet worden dat de tijdvakken, zoals hierboven wordt verduidelijkt, elkaar kunnen overlappen en dat de onderzoeker zich voor het probleem gesteld ziet dat hij of zij aan het ene tijdvak een hogere prioriteit moet toekennen dan aan het andere tijdvak.

In de JLD wordt sinds 2004 bijgehouden wat iemands geschiedenis is wat betreft sociale uitkeringen, arbeidsmarktactivering, opleiding en werk, met ongeveer 13 miljoen uitkerings- en werktijdvakken voor ongeveer twee miljoen personen. Elk tijdvak heeft een start- en einddatum en een operationele code<sup>41</sup> waarmee in elke periode de situatie van een persoon kan worden vastgesteld. In de JLD is een groot aantal "variabelen" opgenomen: gender, leeftijd, burgerlijke staat, staatsburgerschap, opleidingsniveau, vorig beroep, arbeidsverleden en werkloosheidsverleden (duur en aantal tijdvakken), baankenmerken (zoals sector), verleden op het gebied van re-integratieopleidingen (soort, duur en aantal tijdvakken), soort uitkering (JA, JB), aantal kinderen, afhankelijke personen, soort gezinstoelage (d.w.z. toelagen voor meerderjarige en minderjarige afhankelijke personen, alleen meerderjarigen enz.), inkomsten, belastingen en geografische locatie<sup>42</sup>.

Met behulp van individuele identificatiecodes wordt de JLD bovendien gekoppeld aan andere administratieve gegevens. Bij de evaluatie van JobsPlus, bijvoorbeeld, werd de JLD gekoppeld aan een aparte monitoringsdatabank met uitvoerige gegevens over de begin- en einddata van deelname aan JobsPlus, het aantal dagen dat iemand bij aanvang van JobsPlus in het Live Register staat ingeschreven en het soort behandeling.

Er zijn twee belangrijke manieren om inzage in de JLD-gegevens te krijgen: bij contractonderzoek, dat plaatsvindt wanneer het DSP een aanbestedingsprocedure in gang zet; of op verzoek van geaccrediteerde onderzoekers. In dit tweede geval worden de verzoeken om toegang per keer bekeken. Als een verzoek wordt ingewilligd, moet er een juridisch afdwingbare overeenkomst over gegevensuitwisseling worden ondertekend tussen de onderzoekers/instellingen en het DSP. Alle gegevens worden gepseudonimiseerd en alleen de gegevensvelden die relevant zijn voor het onderzoeksthema worden toegezonden. De onderzoeker of de onderzoeksinstelling heeft beperkt toegang tot de gegevens en aan het einde van het onderzoeks- of evaluatieproject moeten de gegevens worden gewist. Samen met de JLD-gegevens krijgen onderzoekers een codeboek, hoewel de documentatie forse verbetering behoeft.

In de loop der jaren is voor uiteenlopende onderzoeksactiviteiten en evaluaties gebruikgemaakt van de JLD. Back to Education Allowance (2015), JobBridge Activation Programme (2016), Back to Work Enterprise Allowance (2017), JobPath (2019) en JobPlus (2020) zijn enkele van de programma's die voor een evaluatie zijn geselecteerd.

De structuur en inhoud van de JLD zijn sinds de beginstadia verbeterd. Het vergt echter voortdurende inzet om de JLD actueel te houden, verbeteringen te bedenken en lacunes te dichten, bijvoorbeeld door uitvoeriger informatie over onderwijs en meer programmaspecifieke gegevens (inhoud, voltooiing enz.) op te nemen en de inkomensgegevens te verbeteren. Er staat een project op stapel om de JLD op de volgende punten te verbeteren: periodieke actualiseringen (bijvoorbeeld maandelijks) door middel van automatisering en het testen van de gegevenspijplijn, en het toevoegen of vervangen van gegevensbronnen om ervoor te zorgen dat alle werk- en opleidingssteun, andere socialezekerheidsprogramma's en realtime inkomensgegevens worden meegenomen<sup>43</sup>.

### **Het instituut voor arbeidsmarkt- en beroepsonderzoek (*Institut für Arbeitsmarkt- und Berufsforschung, IAB*) in Duitsland en de geïntegreerde beroepsbiografieën (*Integrierten Erwerbsbiografien, IEB*)**

De toegang tot gegevens voor wetenschappelijke doeleinden is in Duitsland verbeterd als gevolg van de arbeidsmarkthervormingen van 2003-2005, toen de nadruk kwam te liggen op het evalueren van beleid, en

kunnen maken.

<sup>41</sup> Deze code staat voor een combinatie van activiteiten op het gebied van arbeidsmarktactivering en opleiding, sociale uitkeringen en tijdvakken van arbeid.

<sup>42</sup> Een aspect dat niet onderbelicht mag blijven is dat sommige gegevens betrekking hebben op verschillende perioden. Zo wordt het inkomen bijvoorbeeld op jaarbasis geregistreerd, terwijl de sociale uitkeringen wekelijks worden bijgewerkt.

<sup>43</sup> Wij bedanken Frank Humphreys, Ciaran Judge, Saidhbhín Hardiman en Krzysztof Gigon van het DSP voor de informatie over de JLD.



dankzij de aanbeveling van de Commissie ter verbetering van de informatie-infrastructuur tussen wetenschap en statistiek (*Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik*) om bij elke publieke producent van microgegevens een centrum voor onderzoeksgegevens op te richten. Als gevolg daarvan heeft de federale arbeidsdienst (*Bundesagentur für Arbeit*) in 2004 binnen het instituut voor arbeidsmarkt- en beroepsonderzoek (IAB) een centrum voor onderzoeksgegevens opgericht. Het IAB is verantwoordelijk voor het ophalen van gegevens uit administratieve procedures, zodat datasets kunnen worden samengesteld die nuttig zijn voor empirisch onderzoek.

Het IAB beheert meer dan 15 datasets waarover de wetenschappelijke wereld kan beschikken. De gegevens zijn afkomstig uit administratieve gegevens van het socialezekerheidsstelsel, uit interne processen van de federale arbeidsdienst en uit door het IAB gehouden enquêtes<sup>44</sup>.

Wat het socialezekerheidsstelsel betreft, geldt dat alle werkgevers verplicht zijn om een aantal aspecten en kenmerken van hun werknemers te rapporteren. Deze gegevens vormen een rijke bron van informatie over het arbeidsverleden van mensen. De administratieve gegevens met betrekking tot de interne procedures van de federale arbeidsdienst hebben onder meer betrekking op: de verplichte werkloosheidsverzekering, werkloosheidsuitkeringen en de perioden waarop aanspraak op die uitkeringen kan worden gemaakt, adviesbijeenkomsten met werklozen, aangeboden banen en actieve arbeidsmarktmaatregelen. Het IAB combineert de gegevens tot één uitgebreide dataset, de zogeheten “geïntegreerde beroepsbiografieën” (IEB). In 1975 werd begonnen met het verzamelen van deze administratieve gegevens, hoewel niet alle variabelen voor de gehele waarnemingsperiode beschikbaar zijn. De IEB kunnen worden geïntegreerd met verschillende gegevens die afkomstig zijn uit enquêtes die het IAB heeft uitgevoerd, zoals: het *IAB-Betriebspanel*, de Duitse vacature-enquête van het IAB, de *German Management and Organizational Practices* (GMOP); het *Panel Arbeitsmarkt und soziale Sicherung* (PASS); *Arbeiten und Lernen im Wandel*; de werknemersenquête *Bonuszahlungen, Lohnzuwächse und Gerechtigkeit* (BLoG); *Linked-Employer-Employee-Daten des IAB* (LIAB); het *Panel “WeLL” — Arbeitnehmerbefragung für das Projekt “Berufliche Weiterbildung als Bestandteil Lebenslangen Lernens”*.

Het IAB werkt zijn gegevensproducten regelmatig bij en stelt voor onderzoeksdoeleinden diverse gegevensmonsters van deze uitvoerige administratieve gegevensbronnen beschikbaar. Het IAB biedt voor alle gegevensproducten uitvoerige Duits- en Engelstalige documentatie. De rechtsgrondslag voor gegevenstoegang staat in het Duitse sociale wetboek (meerdere versies). Om precies te zijn is voorzien in vier typen gegevenstoegang voor de wetenschappelijke wereld: *Campus Files* (geheel geanonimiseerd en alleen nuttig voor het onderwijs); *Scientific Use Files* zijn geanonimiseerde microgegevens die aan wetenschappelijke instellingen in Duitsland en in EU-lidstaten worden verstrekt ten behoeve van onderzoeksprojecten op arbeidsmarktgebied, maar die niet zijn bedoeld voor het onderwijs of voor onderzoek waarmee commerciële belangen gemoeid zijn. De wetenschappelijke instelling die de gegevens aanvraagt, moet de beveiliging van de gegevens waarborgen; slecht geanonimiseerde gegevens met uitvoeriger informatie zijn alleen op locatie toegankelijk. Het IAB biedt in Neurenberg en diverse andere locaties in Duitsland gescheiden werkplekken in een beveiligde computeromgeving, evenals op diverse plaatsen in de Verenigde Staten en het Verenigd Koninkrijk. Onderzoekers hebben rechtstreeks toegang tot de gegevens, maar krijgen alleen de output van hun programma nadat de openbaarmaking daarvan door IAB-medewerkers is beoordeeld. Uitvoering op afstand betekent dat onderzoekers hun programma voeden met synthetische gegevens en deze uploaden naar JoSuA (*Job Submission Application*). Tijdens dit procedé zien de onderzoekers nimmer de oorspronkelijke gegevens en krijgen zij alleen de voor hen bestemde resultaten. Voor alle vormen van gegevenstoegang bestaan gestandaardiseerde aanvraagformulieren. Nadat een aanvraag is goedgekeurd, wordt tussen de instelling van de onderzoeker en het IAB een overeenkomst gesloten voor het gebruik van de gegevens voor een bepaald project en binnen een bepaalde termijn. In de overeenkomst worden de gegevensbeschermingsregels vastgelegd; inbreuken daarop worden zwaar gesanctioneerd. Sommige datasets die door het IAB worden beheerd (zoals de gekoppelde datasets), kunnen alleen op locatie worden gebruikt.

Op basis van de door het IAB gegenereerde gegevens wordt het actieve arbeidsmarktbeleid regelmatig geëvalueerd. Daarvoor worden de nieuwste empirische methoden gebruikt. In enkele gevallen waren de

<sup>44</sup> In 2011 is bij het IAB het *German Record Linkage Center* opgericht, een gezamenlijk project van het IAB en de Universiteit Duisburg-Essen dat werd gefinancierd door de *Deutsche Forschungsgemeinschaft*, de Duitse stichting voor wetenschappelijk onderzoek. Dit centrum wil het makkelijker maken om datasets zonder speciale identificatiecode te kunnen koppelen.

bevindingen aanleiding tot een beleidswijziging<sup>45</sup>. Het IAB biedt niet alleen onderzoekers in Duitsland maar ook in andere landen inzage in zijn datasets en het aantal gebruikers neemt gestaag toe (in 2016, bijvoorbeeld, had bijna een derde van alle overeenkomsten inzake gegevensgebruik betrekking op instellingen buiten Duitsland).

Zie voor het gebruik van administratieve gegevens voor CIE's en daarmee samenhangende praktische kwesties tevens de door onderzoekers van het JRC opgestelde Europese gids, [Europese Commissie, 2020](#).

### **Wat zijn de mogelijke problemen op het gebied van gegevensbescherming?**

Het kan lastig zijn om de hand te leggen op gegevens waarmee personen of bedrijven kunnen worden geïdentificeerd die hebben deelgenomen aan uit het ESF+ gefinancierde interventies<sup>46</sup>. Voor CIE's zijn microgegevens nodig: gegevens met waarnemingen over individuele eenheden in zowel de behandelings- als de controlegroep.

In bijlage I bij Verordening (EU) 2021/1057 (hierna de "ESF+-verordening" genoemd) is bepaald dat de gegevens over deelnemers moeten worden uitgesplitst naar gender, arbeidsmarktsituatie, leeftijdsgroep, opleidingsniveau en kwetsbare groep (migranten, minderheden, mensen met een handicap, andere achterstandsgroepen). Bij de GB- en de ESF+-verordening is voor 2021-2027 aan beheerautoriteiten de wettelijke verplichting opgelegd om persoonsgegevens te verzamelen en te verwerken in de vorm van documentatie over individuele deelnemers. Daarnaast is in artikel 17, lid 6, van de ESF+-verordening met betrekking tot monitoring en indicatoren het volgende bepaald: "Wanneer gegevens in registers of gelijkwaardige bronnen beschikbaar zijn, kunnen de lidstaten de beheerautoriteiten en andere instanties die belast zijn met de verzameling van gegevens die nodig zijn voor de monitoring en de evaluatie van de algemene steun van het ESF+-onderdeel in gedeeld beheer, in staat stellen gegevens uit die registers of gelijkwaardige bronnen te verkrijgen, overeenkomstig artikel 6, lid 1, punten c) en e), van Verordening (EU) 2016/679."

*De ESF+-  
verordening over  
gegevensbeheer*

Deze regels, die zijn vastgelegd in de VGB- en de ESF+-verordening, vergemakkelijken de inzage en het gebruik van persoonsgegevens die voor ESF+-monitoring nodig zijn en op grond waarvan in het geval van CIE's de behandelingsgroep wordt samengesteld. De inzage van persoonsgegevens die nodig zijn om controlegroepen samen te stellen alsook de verwerking van gegevens over de behandelings- en controlegroep moeten echter in overeenstemming zijn met Verordening (EU) 2016/679 (de algemene verordening gegevensbescherming, AVG), die betrekking heeft op de algemene doorgifte en het gebruik van persoonsgegevens, waaronder bijzondere gegevenscategorieën<sup>47</sup> binnen de EU. In het volgende tekstkader

<sup>45</sup> Een voorbeeld is de evaluatie van de verplichte arbeidsmarktintegratieovereenkomst tussen de werkzoekende en de dossierbeheerder. Het IAB kon met een gerandomiseerd veldexperiment, en kijkend naar de arbeidsmarktbiografieën van de mensen die meededen aan het experiment, laten zien dat de verplichte regeling bij sommige groepen werklozen averechts uitpakte en dat dit instrument flexibeler moet worden toegepast (van den Berg et al., 2016).

<sup>46</sup> Zie [Ismeri Europa — Ecorys — Institute for Employment Studies, 2019](#)

<sup>47</sup> In de AVG wordt niet langer het begrip "gevoelige gegevens" gebruikt, maar wordt nu gewag gemaakt van "bijzondere gegevenscategorieën". Volgens artikel 9 gaat het daarbij om: "persoonsgegevens waaruit ras of etnische afkomst, politieke opvattingen, religieuze of levensbeschouwelijke overtuigingen, of het lidmaatschap van een vakbond blijken, [...] genetische gegevens, biometrische gegevens met het oog op de unieke identificatie van een persoon, of gegevens over gezondheid, of gegevens met betrekking tot iemands seksueel gedrag of seksuele gerichtheid".

staat een overzicht van de hoofdinhoud en de belangrijkste verplichtingen zoals die door de AVG worden geregeld.

### Tekstkader 8 EU-regelgevingskader voor de verwerking van persoonsgegevens

De relevante wetgeving inzake de verwerking van persoonsgegevens in Europa bestaat in hoofdzaak uit Verordening (EU) 2016/679 en uit de richtsnoeren en maatregelen die door bevoegde autoriteiten zoals het Europees Comité voor gegevensbescherming (EDPB) zijn vastgesteld. Ter bescherming van de rechten en vrijheden van betrokkenen zijn in deze wetgeving tal van voorwaarden en beperkingen inzake de verwerking van persoonsgegevens vastgelegd. In ieder geval moet de noodzaak tot bescherming en waarborging die uit de bepalingen en verordeningen naar voren komt in evenwicht worden gebracht met de noodzaak om te voorkomen dat het wetenschappelijk onderzoek wordt beknot. De noodzaak tot bescherming en waarborging moet juist fungeren als troefkaart bij de ontwikkeling van dat onderzoek. Om die reden heeft de Europese Unie, binnen dit rechtskader, bepaald dat er in specifieke omstandigheden uitzonderingen kunnen worden getroffen ten behoeve van onderzoeksactiviteiten en de verspreiding van outputs, mits het belangrijkste recht van de betrokkenen, namelijk het recht op privacy, wordt gewaarborgd<sup>48</sup>.

#### VERORDENING (EU) 2016/679 (ALGEMENE VERORDENING GEGEVENSBESCHERMING)

De algemene verordening gegevensbescherming (AVG) is op 24 mei 2016 in werking getreden en is op 25 mei 2018 in alle lidstaten volledig van toepassing geworden. De AVG "is van toepassing op de geheel of gedeeltelijk geautomatiseerde verwerking, alsmede op de verwerking van persoonsgegevens die in een bestand zijn opgenomen of die bestemd zijn om daarin te worden opgenomen"<sup>49</sup>.

Wat de territoriale werkingssfeer betreft, is de verordening zowel van toepassing op de verwerking van persoonsgegevens door personen die zich op het grondgebied van de Europese Unie bevinden als op gevallen waarin de verwerking betrokkenen aangaat die zich in de Europese Unie bevinden, ook indien de verwerkingsverantwoordelijke of de verwerker zich buiten de EU bevindt.

Inhoudelijk gezien en vergeleken met de vorige wetgeving zorgt de verordening voor een versterking van de rechten van betrokkenen en legt zij verwerkingsverantwoordelijken een reeks verplichtingen op, waarbij een logica wordt gevolgd die berust op een risicoanalyse en op het verantwoordingsbeginsel. Daarnaast moet de verwerkingsverantwoordelijke volgens de verordening voldoen aan een reeks vereisten die in de vorige wetgeving ontbraken, zoals de gegevensbeschermingseffectbeoordeling (de artikelen 35 en 36), minimale gegevensverwerking overeenkomstig de criteria voor gegevensbescherming door ontwerp en door standaardinstellingen (artikel 25), het opzetten van een register van de verwerkingsactiviteiten (artikel 30) en de aanwijzing van de functionaris voor gegevensbescherming (DPO) (de artikelen 37 tot en met 39).

Het is belangrijk om te wijzen op de hoofdbeginselen die in artikel 5 van de verordening zijn vastgesteld en waaraan de verwerkers van persoonsgegevens moeten voldoen. Het betreft met name:

- a) rechtmatigheid, behoorlijkheid en transparantie: persoonsgegevens moeten worden verwerkt op een wijze die rechtmatig, behoorlijk en transparant is;
- b) doelbinding: persoonsgegevens moeten voor welbepaalde, uitdrukkelijk omschreven en legitieme doeleinden worden verzameld en mogen vervolgens niet verder op een met die doeleinden onverenigbare wijze worden verwerkt. Bij verwerking met het oog op statistische doeleinden of wetenschappelijk onderzoek moet de verwerkingsverantwoordelijke de benodigde garanties en beschermingsmaatregelen treffen;
- c) minimale gegevensverwerking: alleen gegevens die strikt noodzakelijk zijn voor de verwezenlijking van specifieke doeleinden mogen worden verzameld;
- d) juistheid: de verzamelde gegevens moeten juist zijn en zo nodig worden bijgewerkt;
- e) opslagbeperking: persoonsgegevens moeten worden bewaard in een vorm die het mogelijk maakt de

<sup>48</sup> Voorbeelden van andere relevante wetgeving die bijdraagt aan het rechtskader voor de verwerking van persoonsgegevens zijn: Verordening (EU) nr. 557/2013 betreffende de Europese statistiek, wat de toegang tot vertrouwelijke gegevens voor wetenschappelijke doeleinden betreft; Verordening (EU) 2018/1725 betreffende de bescherming van natuurlijke personen in verband met de verwerking van persoonsgegevens door de instellingen, organen en instanties van de Unie en betreffende het vrije verkeer van die gegevens.

<sup>49</sup> Artikel 2, lid 1, AVG.

betrokkenen te identificeren, maar niet langer dan noodzakelijk is voor de verwezenlijking van de doeleinden waarvoor de persoonsgegevens worden verwerkt. Daartoe mogen persoonsgegevens voor langere perioden worden opgeslagen voor zover de gegevens louter met het oog op archivering in het algemeen belang, wetenschappelijk of historisch onderzoek of statistische doeleinden worden verwerkt overeenkomstig artikel 89. In die gevallen moeten ten aanzien van de gegevens passende technische en organisatorische maatregelen worden getroffen om de rechten en vrijheden van de betrokkenen te beschermen;

- f) *integriteit en vertrouwelijkheid*: persoonsgegevens moeten met gebruikmaking van passende technische of organisatorische middelen op een dusdanige manier worden verwerkt dat de beveiliging ervan gewaarborgd is, en dat zij onder meer beschermd zijn tegen ongeoorloofde of onrechtmatige verwerking en tegen onopzettelijk verlies, vernietiging of beschadiging;
- g) *verantwoordingsplicht*: dit is een van de belangrijkste beginselen van de verordening. Hierin is bepaald dat de verwerkingsverantwoordelijke verantwoordelijk is voor de gegevensverwerking en kan aantonen dat de vereiste maatregelen zijn uitgevoerd.

Dit geheel van beginselen vormt de hoofdstructuur waarop de AVG berust en is richtinggevend voor een reeks verplichtingen waaraan de verwerkingsverantwoordelijke en de verwerker moeten voldoen.

## VERWERKING VAN PERSOONSGEGEVENS IN HET KADER VAN CIE'S: VOORWAARDEN, BEPERKINGEN EN BELANGRIJKSTE PROBLEMEN

Bij een contrafeitelijk effectonderzoek worden grote hoeveelheden gegevens verwerkt, waaronder persoonsgegevens. In het licht van het hierboven beschreven regelgevingskader is het belangrijk om op de hoogte te zijn van de voorwaarden, beperkingen en belangrijkste problemen waarmee de lidstaten en de beheerautoriteiten bij het uitvoeren van CIE's worden geconfronteerd.

### **Rechtsgrond en doeleinden van de verwerking**

Voor een rechtmatige verwerking van persoonsgegevens moet aan ten minste één van de in artikel 6 AVG vermelde voorwaarden zijn voldaan<sup>50</sup>. Als de verwerking echter betrekking heeft op "bijzondere gegevenscategorieën", moet ook op artikel 9 AVG worden gewezen.

In het kader van een CIE lijkt het in artikel 6, lid 1, punt e), AVG genoemde **algemeen belang** ("de verwerking is noodzakelijk voor de vervulling van een taak van algemeen belang of van een taak in het kader van de uitoefening van het openbaar gezag dat aan de verwerkingsverantwoordelijke is opgedragen") de meest geschikte rechtsgrond voor het verwerken van gegevens. Zoals bepaald in artikel 6, lid 3, AVG moet deze rechtsgrond bij het nationale of Unierecht worden vastgesteld. Het "algemeen belang" slaat duidelijk op de in artikel 17, lid 6, van Verordening (EU) 2021/1057 (de ESF+-verordening) beschreven verplichtingen van de beheerautoriteiten met betrekking tot het gebruik van "in registers of gelijkwaardige bronnen" beschikbare gegevens. Daarnaast staat in artikel 4 van de GB-verordening: "De lidstaten en de Commissie mogen persoonsgegevens alleen verwerken indien dat nodig is voor het vervullen van hun respectieve verplichtingen uit hoofde van deze verordening, met name voor monitoring, rapportage, communicatie, bekendmaking, evaluatie [...]". Aan verwerkingsverantwoordelijken kunnen bij nationale wetgeving ook soortgelijke bevoegdheden voor het beheren en verwerken van gegevens in het algemeen belang worden toegekend.

Andere rechtsgronden voor een CIE staan in de overige in artikel 6, lid 1, AVG genoemde voorwaarden, in het bijzonder:

- de **toestemming** van de betrokkene (artikel 6, lid 1, punt a), AVG). Deze toestemming kan bijvoorbeeld een geschikte rechtsgrond vormen wanneer met behulp van een enquête gegevens voor het CIE worden verzameld en de betrokkenen eenvoudig toestemming voor verwerking kunnen geven (zie inzake toestemming artikel 7 AVG). Over het algemeen ligt toestemming ingewikkelder indien die niet

<sup>50</sup> Deze voorwaarden zijn: a) de betrokkene heeft toestemming gegeven voor de verwerking van zijn persoonsgegevens voor een of meer specifieke doeleinden; b) de verwerking is noodzakelijk voor de uitvoering van een overeenkomst waarbij de betrokkene partij is, of om op verzoek van de betrokkene vóór de sluiting van een overeenkomst maatregelen te nemen; c) de verwerking is noodzakelijk om te voldoen aan een wettelijke verplichting die op de verwerkingsverantwoordelijke rust; d) de verwerking is noodzakelijk om de vitale belangen van de betrokkene of van een andere natuurlijke persoon te beschermen; e) de verwerking is noodzakelijk voor de vervulling van een taak van algemeen belang of van een taak in het kader van de uitoefening van het openbaar gezag dat aan de verwerkingsverantwoordelijke is opgedragen; f) de verwerking is noodzakelijk voor de behartiging van de gerechtvaardigde belangen van de verwerkingsverantwoordelijke of van een derde, behalve wanneer de belangen of de grondrechten en de fundamentele vrijheden van de betrokkene die tot bescherming van persoonsgegevens nopen, zwaarder wegen dan die belangen, met name wanneer de betrokkene een kind is.

ruim van tevoren wordt gepland. Indien de andere rechtsgronden niet van toepassing zijn, kan toestemming als “residuele” rechtsgrond voor een CIE worden beschouwd;

- de verwerking is noodzakelijk om te **voldoen aan een wettelijke verplichting** die op de verwerkingsverantwoordelijke rust (artikel 6, lid 1, punt c), AVG, waarnaar verwezen wordt in artikel 17 van de ESF+-verordening). Deze rechtsgrond moet bij wet worden vastgesteld en kan de betrokkenheid van particuliere of publieke instanties inhouden. Vanwege bepaalde wettelijke voorschriften kan het bijvoorbeeld zo zijn dat een voor een dataset verantwoordelijke particuliere of publieke instantie in het kader van een CIE moet samenwerken met de beheerautoriteit.

Daarnaast kan de verwerkingsverantwoordelijke **de verzamelde gegevens voor andere doeleinden gebruiken indien die verenigbaar zijn met de oorspronkelijke doeleinden**. De verwerkingsverantwoordelijke moet dit toetsen aan de voorwaarden uit artikel 6, lid 4, en artikel 5, lid 1, punt b), AVG. Deze bepalingen zijn met name relevant bij de verwerking van gegevens met het oog op wetenschappelijk onderzoek of voor statistische doeleinden. Op grond daarvan is het mogelijk om administratieve gegevens te gebruiken voor andere doeleinden dan de oorspronkelijke doeleinden. Voor die andere doeleinden is geen specifieke toestemming nodig, maar zij moeten wel voldoen aan de waarborgen van artikel 89, lid 1, en met name wat pseudonimisering betreft (zie hieronder). Deze bepalingen kunnen bijvoorbeeld relevant zijn wanneer voor een CIE gegevens uit een werklozenregister worden gebruikt.

Vanwege het specifieke karakter van CIE's hoeft de verwerkingsverantwoordelijke die de gegevens verzamelt niet de entiteit te zijn die het onderzoek uitvoert. Er moet dan sprake zijn van omstandigheden waaronder de doorgifte van gegevens gerechtvaardigd is en waardoor de ontvanger verder kan gaan met het CIE.

#### ***Anonimisering en pseudonimisering: verwerking voor statistische doeleinden***

Volgens artikel 89 AVG moet de verwerking van gegevens met het oog op het algemeen belang of in het kader van wetenschappelijk onderzoek of voor statistische doeleinden onderworpen zijn aan passende waarborgen voor de rechten en vrijheden van de betrokkene en moet bovenal het beginsel van “minimale gegevensverwerking” in acht worden genomen. Dit betekent dat er pseudonimiseringstechnieken moeten worden toegepast<sup>51</sup>.

Wanneer die doeleinden kunnen worden verwezenlijkt door verdere verwerking op basis waarvan de betrokkenen niet of niet meer geïdentificeerd kunnen worden, moeten die doeleinden worden verwezenlijkt door de gegevens te anonimiseren en die gegevens vervolgens in samengevoegde vorm te verwerken. Bij CIE's worden resultaten per definitie samengevoegd en doet dit risico zich niet voor, behalve wanneer de oorspronkelijke datasets om wetenschappelijke redenen worden gepubliceerd. In dat geval moeten de datasets worden geanonimiseerd (zie het voorbeeld in het onderstaande tekstkader).

#### ***Gegevensopslag en beveiligde verwerking***

Gegevensopslag is een van de belangrijkste aspecten van gegevensverwerking. In de wetgeving is niet gepreciseerd hoe gegevens moeten worden opgeslagen, maar volgens de bovengenoemde beginselen moeten de opslag en verwerking altijd verband houden met het doel van het onderzoek. Als het doel van de verwerking is verwezenlijkt en de gegevens niet langer hoeven te worden bewaard, moet de gegevensbewaring worden gestaakt. Deze vuistregel moet worden opgenomen in de privacyverklaring die aan betrokkenen wordt verstrekt wanneer zij in administratieve datasets worden opgenomen. Omdat het waarschijnlijk ondoenlijk is alle betrokkenen persoonlijk te informeren, kunnen ook alternatieve wijzen van informatieverschaffing worden gebruikt (op de website van de beheerautoriteit kan bijvoorbeeld een informatiepagina worden aangemaakt met het privacybeleid inzake onderzoeksactiviteiten).

Meer in het bijzonder is het zo dat het doel geacht wordt te zijn verwezenlijkt zodra het CIE is uitgevoerd, de resultaten beschikbaar zijn gemaakt en deze voor statistisch gebruik zijn samengevoegd. Op dat moment moeten de gegevens ofwel worden geanonimiseerd (voor het geval ze in een later stadium opnieuw en voor andere doeleinden worden gebruikt) ofwel worden verwijderd. Voor zover dat bij wet- en

<sup>51</sup> Pseudonimisering is de verwerking van persoonsgegevens op zodanige wijze dat deze niet meer aan een specifieke persoon kunnen worden gekoppeld zonder dat er aanvullende gegevens worden gebruikt. Anonimisering houdt in dat persoonsgegevens op zodanige wijze worden verwerkt dat het met die gegevens onmogelijk is geworden om personen te identificeren. Zie voor een overzicht van pseudonimiseringstechnieken: <https://www.enisa.europa.eu/publications/pseudonymisation-techniques-and-best-practices>

regelgeving is toegestaan, kunnen de gegevens voor onbepaalde tijd worden opgeslagen, mits uitdrukkelijk naar de regelgeving wordt verwezen.

De andere voorwaarde die in de AVG wordt gesteld is dat de verwerking moet zijn ingebed in toereikende en passende beveiligingsmaatregelen. Volgens artikel 32 van de verordening moet degene die de verwerking uitvoert (de verwerkingsverantwoordelijke of de verwerker) passende technische en organisatorische maatregelen treffen om een op het risico afgestemd beveiligingsniveau te waarborgen. Dit betekent dat er geen standaardmaatregelen zijn voorgeschreven, maar dat de maatregelen per geval moeten worden vastgesteld en dat de risico's die de afzonderlijke verwerking oplevert specifiek moeten worden vermeld.

## WETTELIJKE VERPLICHTINGEN

### ***Informatie voor de betrokkenen***

De eerste taak van de verwerkingsverantwoordelijke is de informatieverstrekking aan de betrokkenen. Deze verplichting is vastgelegd in artikel 13 AVG en geldt wanneer persoonsgegevens bij de betrokkene worden verzameld. Is dat niet het geval, dan geldt artikel 14 als uitgangspunt. De betrokkene moet worden ingelicht over de doeleinden en methoden van de verwerking, de rechtsgrond, de opslagtermijnen en de rechten die de betrokkene kan uitoefenen. Bij CIE's wordt deze informatie verstrekt op het moment waarop de gegevens worden verzameld die later voor statistische en evaluatiedoeleinden zullen worden gebruikt, zoals voor datasets in verband met ESF-monitoring of datasets uit het werklozenregister. Als die informatieverstrekking praktisch gezien niet mogelijk is, zoals hierboven is aangegeven, en vooral wanneer het gaat om personen in registers die voor het CIE worden gebruikt, kan gebruik worden gemaakt van een website of van algemene informatiemiddelen.

### ***Governance van de verhouding tussen de verschillende entiteiten die bij de verwerking betrokken zijn***

Bij onderzoeksactiviteiten zijn over het algemeen meerdere instanties betrokken. De onderlinge verhoudingen moeten in die gevallen worden geregeld in specifieke regelingen die bekend staan als "gegevensverwerkingsovereenkomsten"<sup>52</sup>. Deze moeten per geval worden vastgesteld en laten zien wat de bijdrage van de diverse belanghebbenden is. Voorbeelden van mogelijke verhoudingen zijn:

- *verwerkingsverantwoordelijke — verwerkingsverantwoordelijke*: hiervan is sprake wanneer de entiteiten samenwerken bij de uitvoering van een project — zij het onder verschillende omstandigheden en met verschillende taken — waarbij elke entiteit haar eigen onderscheidende verwerkingsdoel behoudt;
- *verwerkingsverantwoordelijke — verwerker*: hiervan is sprake wanneer één entiteit (de verwerkingsverantwoordelijke) het doel of de middelen van de verwerking bepaalt en voor bepaalde verwerkingsactiviteiten een andere entiteit (de verwerker) inschakelt. In dat geval wordt verwezen naar de voorwaarden en verplichtingen die zijn vastgesteld in artikel 28 AVG;
- *gezamenlijke verwerkingsverantwoordelijken*: deze verhouding wordt beschreven in artikel 26 AVG, waarin is bepaald: "Wanneer twee of meer verwerkingsverantwoordelijken gezamenlijk de doeleinden en middelen van de verwerking bepalen, zijn zij gezamenlijke verwerkingsverantwoordelijken." In die gevallen moeten de partijen op transparante wijze hun respectieve verantwoordelijkheden bepalen voor de naleving van de regelgeving, met name wat de rechten van de betrokkenen betreft.

### ***Gegevensbeschermingseffectbeoordeling***

Wanneer een verwerking van persoonsgegevens een hoog risico inhoudt voor de rechten en vrijheden van betrokkenen moet de verwerkingsverantwoordelijke volgens artikel 35 AVG vóór de verwerking een effectbeoordeling<sup>53</sup> uitvoeren. Bij CIE's is een effectbeoordeling alleen van belang indien er op grote schaal bijzondere categorieën betrokkenen worden verwerkt. In zulke gevallen kan de verwerkingsverantwoordelijke in de overheidsdienst een effectbeoordeling uitvoeren overeenkomstig de voorschriften van de nationale gegevensbeschermingsautoriteit en met de instrumenten die door die autoriteit ter beschikking zijn gesteld.

<sup>52</sup> Hiervan is bijvoorbeeld sprake wanneer de verwerkingsverantwoordelijke niet degene is die het onderzoek daadwerkelijk uitvoert. In zo'n geval moet een overeenkomst worden gesloten waarin de verhouding tussen de partijen wordt geregeld wat betreft de toegang tot gegevens en de uitvoering van het onderzoek.

<sup>53</sup> Zie voor meer informatie de richtsnoeren die door het EDPB zijn uitgegeven <https://ec.europa.eu/newsroom/article29/items/611236>

**TOEKOMSTPERSPECTIEVEN: DE DATAGOVERNANCEVERORDENING**

Het hierboven geschetste kader geeft de huidige stand van zaken weer en de regels die thans in de EU van kracht zijn. De AVG werd speciaal ingevoerd om een doorbraak op het gebied van de bescherming van persoonsgegevens mogelijk te maken en om in de hele EU een gemeenschappelijk systeem op te zetten dat aansluit bij de huidige technologische ontwikkelingen en de hedendaagse datasamenleving. Er staan verdere hervormingen op stapel die grote gevolgen kunnen hebben voor de wijze waarop in de komende jaren persoonsgegevens worden verzameld.

De datagovernanceverordening verkeert bij de Europese Commissie nog in de fase van voorstel<sup>54</sup>. Deze verordening is uitdrukkelijk bedoeld om de beschikbaarheid van bruikbare gegevens te bevorderen door het vertrouwen in gegevensbemiddelaars te vergroten en de mechanismen voor gegevensuitwisseling in de EU te verbeteren. Persoonsgegevens zullen waarschijnlijk onder de nieuwe norm vallen. Het gebruik van die gegevens voor statistische en onderzoeksdoeleinden kan een hogere vlucht nemen omdat er een nieuwe figuur wordt geïntroduceerd: de bemiddelaar voor het delen van persoonsgegevens. Die stap zou kunnen bijdragen tot het ontstaan van een klimaat dat gunstiger is voor CIE's omdat het makkelijker wordt om op grote schaal grote hoeveelheden informatie in te zien.

Het is duidelijk dat de voorschriften van de AVG geen beletsel vormen voor het uitvoeren van CIE's, maar met het oog op een veilige CIE moeten in alle lidstaten wel enkele procedurele en operationele basisstappen in acht worden genomen. Dit zijn bijvoorbeeld de belangrijkste stappen in een reguliere situatie waarin monitoringgegevens (behandelde personen) worden gecombineerd met gegevens uit een openbaar register (controlegroep) en het CIE wordt uitgevoerd door een externe beoordelaar:

*Belangrijkste stappen voor een betere naleving van de AVG*

1. De beheerautoriteit bereikt overeenstemming met de autoriteiten die verantwoordelijk zijn voor de gegevens (bv. een werklozenregister, belastingregister enz.) die nodig zijn om de controlegroep samen te stellen en de behandelings- en controlegroep te analyseren. Indien er geen uitdrukkelijke toestemming is verleend, controleert de beheerautoriteit of het gebruik van deze gegevens in overeenstemming is met artikel 6, lid 1, punt e), of met artikel 6, lid 4, en artikel 5, lid 1, punt b), AVG.
2. De beheerautoriteit sluit overeenkomsten met de andere entiteiten (de geveenseigenaren en de beoordelaar) waarin de informatiestromen en de wederzijdse verantwoordelijkheden worden geregeld conform de voorschriften van de AVG. De dienstovereenkomst tussen de beoordelaar en de beheerautoriteit moet een specifieke bepaling over de bescherming van gegevens bevatten; in het geval van andere autoriteiten kan de gegevensbescherming bij de uitvoering van CIE's worden geregeld in een memorandum van overeenstemming of in speciale nationale procedures.
3. Op basis van de bovengenoemde overeenkomst ontvangt de beheerautoriteit de gegevens in gepseudonimiseerde vorm van de geveenseigenaar en geeft de beheerautoriteit met het oog op de verwerking van de behandelings- en controlegroep de gegevens overeenkomstig de AVG door aan de beoordelaar<sup>55</sup>.
4. Zoals in de AVG is bepaald, beantwoordt de opslag van persoonsgegevens van de behandelings- en controlegroep uitsluitend aan de voorschriften inzake gegevensopslag voor de duur en de doeleinden

<sup>54</sup> Het voorstel voor een datagovernanceverordening is eind 2020 door de Commissie vastgesteld. Deze website toont de stand van zaken met betrekking tot de datagovernanceverordening: <https://eur-lex.europa.eu/legal-content/NL/HIS/?uri=CELEX%3A52020PC0767>

<sup>55</sup> Indien onderzoek is gedaan naar de behandelings- en controlegroep in de vorm van een enquête, moet er toestemming zijn verleend om de gegevens van de ondervraagde personen voor onderzoeksdoeleinden te gebruiken.

van het onderzoek, onder naleving van de basisvoorschriften voor gegevensbeveiliging. Deze voorschriften worden door alle bij het CIE betrokken entiteiten nageleefd.

De nationale praktijken zijn over het algemeen conform de AVG maar verschillen per lidstaat. Beoordelaars hebben laten weten dat nationale voorschriften inzake gegevensbescherming nog steeds grote belemmeringen voor het gebruik van microgegevens opwerpen. Dit kan komen omdat het tijd vergt om de nationale voorschriften en gebruiken af te stemmen op de AVG, die van recentere datum is, of omdat de AVG door verschillende nationale overheden op verschillende manieren wordt geïnterpreteerd. Het kan ook een gevolg zijn van overregulering, waardoor de AVG in sommige landen strikter wordt toegepast, of van andere onjuiste interpretaties. Om deze belemmeringen in de geest van de AVG en van de ESF+-verordening uit de weg te kunnen ruimen, is in sommige landen een speciaal initiatief van de beheerautoriteiten of de nationale autoriteiten nodig. In het onderstaande tekstkader wordt uitgelegd hoe de Italiaanse regio Veneto enkele van de meest voorkomende problemen heeft aangepakt. Het verdient aanbeveling om contact op te nemen met de nationale gegevensbeschermingsautoriteit en de voorgestelde regelingen te bespreken alvorens die definitief te maken.

### Tekstkader 9 Bescherming en uitwisseling van gegevens

In de “Pilot and feasibility study on the sustainability and effectiveness of results for European Social Fund participants using counterfactual impact evaluations”<sup>56</sup> die in 2019 voor DG EMPL is uitgevoerd, staat een goed voorbeeld van toegang tot geanonimiseerde persoonsgegevens die op betrekkelijk korte termijn werd verkregen.

Die evaluatie betrof een onderzoek van ESF-interventies (investeringsprioriteit 9.i) in vier casussen. Een daarvan werd in 2015/2016 in de regio Veneto uitgevoerd en was op langdurig werklozen gericht. De monitoringgegevens over de deelnemers en de soorten maatregelen werden gecombineerd met gegevens over perioden voor en na de ESF-interventies waarin mensen een baan hadden (“Comunicazioni obbligatorie” geheten). Het proces duurde ongeveer twee maanden (van eind juli tot eind september), gerekend vanaf de eerste bijeenkomst over de gegevensvereisten met de beheerautoriteit en de instantie die over de gegevens beschikte, de regionale instelling Veneto Lavoro. Veneto Lavoro was verantwoordelijk voor het anonimiseren van de gegevens. Dankzij een unieke gemeenschappelijke identificatiecode van Veneto Lavoro konden alle ter beschikking gestelde datasets eenvoudig aan elkaar worden gekoppeld<sup>57</sup>.

De opdracht tot de evaluatie werd weliswaar niet rechtstreeks door de betrokken organisaties gegeven, maar de positieve ervaring is wel te danken aan twee belangrijke factoren: intensieve (zij het informele) samenwerking tussen de regionale kantoren van de beheerautoriteit en de regionale instelling Veneto Lavoro, noodzakelijk voor het aanpakken en oplossen van problemen op het gebied van gegevensbescherming; en het bestaan van een door Veneto Lavoro beheerde databank (Mercurio) voor publiek gebruik die voor onderzoeksdoeleinden ter beschikking wordt gesteld<sup>58</sup>.

Mercurio is een statistische databank met alle bijzonderheden die door de arbeidsbureaus in de regio Veneto worden geregistreerd over de perioden waarin de inwoners een baan hebben of werkloos zijn. Bovendien bevat de databank uitgebreide informatie over alle geregistreerde werknemers en ondernemingen. Ten tijde van de analyse bevatte Mercurio gegevens over meer dan 4 miljoen werknemers, 17,6 miljoen tijdvakken van arbeid en 4 miljoen tijdvakken van werkloosheid<sup>59</sup>.

<sup>56</sup> [Ismeri Europa — Ecorys — Institute for Employment Studies, 2019](#)

<sup>57</sup> Er was een extra dataset beschikbaar met de start- en einddata van alle door de behandelde en niet-behandelde personen ondertekende “verklaringen inzake directe arbeidsinzetbaarheid”.

<sup>58</sup> Zie <https://www.venetolavoro.it/public-use-file> voor een beschrijving van Mercurio en van de regels inzake de toegang daartoe.

<sup>59</sup> De volgende ter zake doende variabelen werden uit de administratieve gegevens geëxtraheerd: gegevens over het begin en het einde van elke periode van arbeid en werkloosheid; het soort arbeidsovereenkomst voor elke gewerkte periode (vast, tijdelijk, stageovereenkomst); verwachte duur van de tijdelijke arbeidsovereenkomsten; sociaal-demografische informatie (gender,



De dataset wordt van tijd tot tijd “bijgewerkt en opgeschoond” en daardoor kon deze vrij eenvoudig worden gematcht met de monitoringgegevens van de beheerautoriteit. Een mogelijke beperking was gelegen in het feit dat het bestand voor publiek gebruik niet vaak genoeg wordt bijgewerkt: om kwalitatief goede gegevens te kunnen garanderen zijn enorme inspanningen nodig<sup>60</sup> (de recentste versie die momenteel beschikbaar is, is bijvoorbeeld in december 2020 bijgewerkt). De ervaring wijst uit dat de beheerautoriteit bij het opstellen van evaluatieplannen voorbereidingen moet treffen voor mogelijke vormen van samenwerking met externe instanties die de voor een contrafeitelijke exercitie benodigde gegevens beheren; oplossingen voor eventuele juridische belemmeringen moet bedenken, en moet aangeven hoe de inzage, het beheer en de verstrekking van administratieve gegevens in geanonimiseerde vorm verloopt.

### 2.3.2. Hoe moet de “behandelingsgroep” worden bepaald?

Om een CIE te kunnen uitvoeren is het essentieel dat vaststaat wat onder “behandeling” of onder “deelname aan de interventie” wordt verstaan. Zodra ondubbelzinnig vast is komen te staan wanneer een persoon of onderneming “behandeld” is, is het bovendien belangrijk dat die personen of ondernemingen kunnen worden geïdentificeerd. Hier worden de voornaamste aspecten met betrekking tot het bepalen van de behandelings- en controlegroep gepresenteerd. In hoofdstuk 3 staat een uitvoeriger toelichting op de methodologische kenmerken van die groepen.

Op het eerste gezicht lijkt het wellicht eenvoudig om vast te stellen wat “deelname” inhoudt. Er zijn echter een aantal aspecten die wellicht niet direct in het oog springen, maar die wel cruciaal zijn en waar grondig over moet worden nagedacht. Worden stagiairs die bijvoorbeeld een opleidingsprogramma volgen en afhaken bij de interventie aangemerkt als “behandeld”? Hoeveel sessies van een opleidingscursus moeten de cursisten hebben bijgewoond voordat zij als deelnemer worden aangemerkt? Ook mogen de verwachtingseffecten niet uit het oog worden verloren. Vooruitlopend op de interventie kunnen sommige uitkeringsontvangers hun uitkering laten stopzetten met de bedoeling om onder activeringsmaatregelen uit te komen. Worden deze personen ook behandeld wanneer zij bijvoorbeeld nooit verschijnen op afspraken bij het arbeidsbureau?

*De behandelingsgroep vaststellen*

Bij het vaststellen van de “behandelingsgroep” wordt een onderscheid gemaakt tussen “intention to treat” en “treatment on the treated”. Beleidsmatig gezien is de kernvraag meestal of er belangstelling bestaat voor de effecten van de geboden mogelijkheid tot deelname aan een interventie of voor de effecten van de deelname zelf. In het eerste geval kunnen degenen aan wie de interventie wordt aangeboden al dan niet deelnemen. In het tweede geval gaat de belangstelling uit naar het effect van de behandeling op de behandel eenheden en bestaat de behandelingsgroep uitsluitend uit deelnemers<sup>61</sup>.

*“Intention to treat” en “treatment on the treated”*

In eerste instantie willen beleidsmakers graag weten wat de netto-effecten van een behandeling op de deelnemers zijn. Bij nader inzien blijkt vaak dat de zaak niet zo eenvoudig ligt. Als kan worden vastgesteld wie een behandelingsaanbod krijgt, is het beleidsmatig gezien wellicht nuttiger om die personen als de behandelingsgroep te beschouwen. Dat geldt met name voor

*Aanbod of daadwerkelijke behandeling*

geboortedatum, onderwijs, staatsburgerschap).

<sup>60</sup> Dit houdt in dat moet worden gewerkt met informatie die actueler is dan de informatie in de recentste versie van Mercurio en dat er extra werkzaamheden nodig zijn om de jongste gegevens over gewerkte perioden te integreren en op te schonen.

<sup>61</sup> Als deelname aan de interventie verplicht is, bestaat het statusonderscheid in wezen niet. Iedereen die een behandeling krijgt aangeboden, moet dan deelnemen. In de meeste gevallen zijn interventies echter niet verplicht (zoals in deze leidraad steeds wordt verondersteld).

situaties waarin deelname aan een interventie niet verplicht is. Beleidsmakers kunnen de mensen aan wie een interventie is aangeboden niet dwingen tot deelname, en dus draait het om de vraag: wat is het effect op de arbeidsparticipatie en het inkomen van mensen die de kans kregen om deel te nemen aan een opleidingsprogramma?

Om de effecten van het behandelingsaanbod op uiteenlopende resultaten te kunnen meten, moet worden vastgesteld aan wie dat aanbod is gedaan. Dat kan in veel gevallen lastig zijn.

### **Vindplaatsen van geschikte gegevens**

Zodra is bepaald wie er zullen worden behandeld en waaruit de behandeling zal bestaan, is het belangrijk om na te denken over de vraag op welke wijze de te behandelen personen met het oog op de evaluatie zullen worden geïdentificeerd. Dit betekent onherroepelijk dat er een gegevensbron moet worden gevonden met personen dan wel ondernemingen die allemaal in de lijst van behandel eenheden worden opgenomen of waaruit een steekproef wordt getrokken. Deze gegevens worden doorgaans uit de monitoringsystemen van het ESF gehaald en uit eventuele andere gegevensbestanden die voor de interventie in kwestie zijn aangemaakt.

*Zoeken naar gegevensbronnen voor behandelde personen*

Vanwege de monitoring- en rapportageverplichtingen van het ESF (en het ESF+) moeten de begunstigde organisaties registreren hoeveel mensen in het kader van een interventie diensten ontvangen. Ook moeten enkele persoonlijke kenmerken van die mensen worden geregistreerd. Bij CIE's moet de interventie een stap verder gaan en microgegevens over de deelnemers opleveren. De beoordelaars hebben voor elke behandel eenheid (onderneming of persoon) een dossier nodig met gegevens over de belangrijkste kenmerken (geslacht, leeftijd, opleidingsniveau enz.). Omwille van de privacy kunnen deze gegevens worden geanonimiseerd/gepseudonimiseerd, maar als voor het CIE een enquête benodigd is, moet worden nagegaan of de AVG-voorwaarden het toelaten om de identiteit van deze eenheden (namen, adressen, telefoonnummers enz.) door te geven zodat zij in een steekproef kunnen worden opgenomen. Verder heeft elke eenheid een eigen unieke identificatiecode nodig, zodat gegevens uit diverse bronnen makkelijker aan elkaar gekoppeld kunnen worden.

### **2.3.3. Aandachtspunten bij het vaststellen van een controlegroep**

Om een schatting van de contrafeitelijke situatie kunnen maken, moet vaak een controlegroep worden aangewezen. Een beperkende factor bij het samenstellen van een controlegroep is de vraag of de interventie voor de deelnemers al dan niet verplicht is en of de interventie binnen een bepaald rechtsgebied generiek wordt uitgevoerd dan wel beperkt blijft tot een bepaald gebied of een beperkte periode. Bij het samenstellen van een geschikte controlegroep spelen aspecten van drieërlei aard: 1) analytische; 2) beleidsmatige; en 3) praktische aspecten.

*Drie aandachtspunten bij het vaststellen van een controlegroep*

### **Vaststelling van een controlegroep vanuit analytisch perspectief**

Een CIE is bedoeld om een zuivere raming te verkrijgen van de effecten van een interventie op uiteenlopende resultaten. Met het oog daarop moet een raming van de contrafeitelijke resultaten worden gemaakt. De ramingen van de contrafeitelijke resultaten worden uit een controlegroep verkregen (zie

punt 1.1). Zoals blijkt uit figuur 6 en figuur 7 wordt een bepaald effect geraamd door het waargenomen resultaat voor de “behandelingsgroep” te verminderen met de raming van het contrafeitelijke resultaat. In hoeverre een effect vertekend is, hangt af van de mate waarin het voor de controlegroep berekende contrafeitelijke resultaat, ceteris paribus, correspondeert met het resultaat voor de behandelingsgroep indien de personen in die groep niet zouden zijn behandeld.

Om zonder randomisatie tot een vergelijkbare controlegroep te komen, moet deze controlegroep op alle wezenlijke punten vergelijkbaar zijn met de behandelingsgroep, zowel wat waarneembare als niet-waarneembare aspecten betreft. Dit geldt voor alle quasi-experimentele methoden en vormt bijgevolg een noodzakelijke voorwaarde voor alle niet-gerandomiseerde CIE's.

Voor bijna alle ESF-interventies geldt a) dat de deelname vrijwillig is (de doelgroep hoeft niet mee te doen aan een interventie); en/of b) dat er sprake is van andersoortige beperkingen. Bij experimentele interventies of instrumenten die beperkt blijven tot een bepaald (rechts)gebied, beschikken de beoordelaars over een reservoir aan eenheden die voor de controlegroep kunnen worden ingezet. Op dit eventuele reservoir moeten bepaalde filters worden toegepast zodat een definitieve selectie overblijft van controle-eenheden die een goede match met de deelnemers (de behandelingsgroep) zijn. In veel gevallen zijn er vier mogelijkheden:<sup>62</sup>

*Opties voor het kiezen van controlegroepen*

- **locatie** — de controle-eenheden lijken op de deelnemers aan de interventie maar bevinden zich in gebieden van de lidstaat waar de interventie niet beschikbaar is (voor zover die gebieden er zijn). Wanneer die controlegroepen en de juiste gegevens beschikbaar zijn, geniet de difference-in-differencesmethode vaak de voorkeur. De inwoners van verschillende regio's kunnen soms sterk op elkaar lijken en die groepen hebben nog niet de kans gehad om deel te nemen aan de interventie of om de uitnodiging daarvoor af te slaan. Daardoor ontbreekt een belangrijke oorzaak van eventuele vertekening. De inwoners van verschillende regio's hebben echter te maken met uiteenlopende arbeidsmarktsituaties. Bij dergelijke variaties presteren de difference-in-differencestechnieken heel behoorlijk, want verschillen in arbeidsmarktsituatie zijn door de tijd heen redelijk constant. Als de effecten worden geraamd door middel van matching, is het echter niet erg raadzaam om voor de controlegroep steekproeven uit verschillende lokale arbeidsmarkten te nemen. Wanneer uit verschillende lokale arbeidsmarkten steekproeven voor de controlegroep worden genomen, kan de vertekening die daardoor ontstaat groter zijn dan de selectievertekening<sup>63</sup>;
- **tijd** — de controle-eenheden lijken op de deelnemers aan de interventie, maar worden op verschillende momenten geobserveerd, hetzij vóór dan wel na de interventie. Wanneer een interventie generiek en verplicht is, moet de controlegroep vaak op deze manier worden gekozen. Dat wil dus zeggen dat alle leden van de doelgroep verplicht moeten deelnemen en dat het programma in het gehele rechtsgebied wordt uitgevoerd. Aan controlegroepen die op die wijze worden samengesteld, kleeft een

<sup>62</sup> Dit gedeelte is gebaseerd op [Card, D., Ibarrran, P. en Villa, J.M., 2011](#)

<sup>63</sup> Zie [Heckman, J.J., Ichimura, H., Smith, J. en Todd, P., 1998](#)

belangrijk nadeel, namelijk dat de resultaten voor die groepen op andere momenten worden gemeten dan de resultaten voor de behandelingsgroep. Dat maakt die resultaten vatbaar voor cyclische schommelingen, wijzigingen in de samenstelling en veranderende macro-economische trends, waardoor het moeilijker kan zijn om een zuiver contrafeitelijk resultaat waar te nemen. Dergelijke controle-eenheden dienen alleen te worden overwogen als er sprake is van resultaten die door de tijd heen weinig variatie vertonen en er geen controlegroep uit eenzelfde periode voorhanden is;

- **voorwaarden** — hierbij worden de controle-eenheden geselecteerd uit qua tijd en plaats gelijke groepen, maar wordt gekozen uit kandidaten die om de een of andere reden niet voor deelname in aanmerking kwamen. Dit type controle-eenheden is vaak gewild wanneer de interventie generiek is, de participatiegraad hoog ligt of deelname verplicht is en er duidelijke selectiecriteria bestaan, zodat personen die “net niet in aanmerking komen” kandidaat voor de controlegroep kunnen zijn. De bedoeling is om groepen te vinden die lijken op de behandelingsgroep, maar die om bekende en vaststaande redenen (die met de gegevens kunnen worden gekwantificeerd) niet voor de behandeling in aanmerking kwamen. De toegang tot interventies in het kader van ESF+ is vaak gebaseerd op duidelijke en makkelijk meetbare selectiecriteria die niet kunnen worden gemanipuleerd (een bepaalde leeftijd van de deelnemers, de duur van iemands werkloosheidsstatus enz.). Bij het kiezen van controle-eenheden kunnen zulke drempelwaarden worden gebruikt als onderscheidende factoren waaromheen de personen in de behandelings- en de controlegroep worden verdeeld;
- **keuze/besef** — feitelijk hebben de behandelingsgroep en de controlegroep (en niet louter de behandelingsgroep) te maken met selectieprocedures op basis van keuzen die zijn ingegeven door factoren die mogelijk niet worden waargenomen<sup>64</sup>. De controle-eenheden kunnen worden gekozen uit personen die wel in aanmerking kwamen maar niet hebben deelgenomen. Dit heeft als voordeel dat die personen veelal afkomstig zijn uit dezelfde arbeidsmarkt als de personen die de behandeling wel hebben ondergaan. Over dergelijke controle-eenheden moet zorgvuldig worden nagedacht, vooral wanneer een op matching gebaseerd CIE-model wordt gebruikt en bij het selectiebesluit uit uitvoerige gegevens kan worden geput. In andere situaties, bijvoorbeeld wanneer de difference-in-differencesmethode wordt toegepast, zijn controle-eenheden met keuze/besef minder aantrekkelijk.

Nog één ander punt moet worden aangestipt. Als er voor zowel de behandelings- als de controlegroep resultaatmetingen van vóór de behandeling beschikbaar zijn, is het belangrijk om voor beide groepen de trends in die resultaatmetingen te onderzoeken. Door de hypothese van de gemeenschappelijke trend te toetsen, wordt het probleem ondervangen van tijdelijke dalingen van participatiegraad en inkomen onder sommige mensen die in aanmerking komen voor actieve arbeidsmarktmaatregelen (anders zouden zij niet voor steun in aanmerking komen, de zogeheten problematiek van de “Ashenfelter Dip”). De beoordelaar zoekt in de resultaatmetingen voor de behandelings- en de controlegroep naar dezelfde temporele trends, zodat

*Onderzoek naar trends van vóór de interventie*

<sup>64</sup> Dit wordt door [Card, D., Ibarrran, P. en Villa, J.M., 2011](#) “tweevoudige selectievertekening” genoemd.

het herstel na een tijdelijke baan of na inkomensverlies niet wordt verward met de relatieve langetermijnopbrengsten die met een CIE moeten worden opgespoord.

Het kiezen van geschikte controlegroepen is een technisch en methodologisch complexe opgave. Aanbevolen wordt dat ambtenaren zich tijdens het opstellen van evaluatieschema's vertrouwd maken met de belangrijkste concepten en vroegtijdig maatregelen nemen om de controle-eenheden te bepalen. Het is belangrijk dat de opdrachtgevers van een evaluatie in een vroege ontwerpfase deskundigen inschakelen die hen van ondersteuning en advies kunnen voorzien.

### Welke beleidsmatige overwegingen zijn relevant?

Het samenstellen van een geschikte controlegroep is niet louter een technische of analytische procedure. Hoewel de analytische aspecten omtrent het bepalen van geschikte controle-eenheden van cruciaal belang zijn, is het in beleidsmatig opzicht ook belangrijk dat een controlegroep een adequaat alternatief voor de interventie vormt.

*Een alternatief voor een interventie zoeken*

CIE's kunnen verschillende vormen aannemen. Met een CIE kunnen bijvoorbeeld de resultaten voor één of meer behandelingsgroepen worden vergeleken met een controlegroep die geen behandeling heeft ondergaan of kunnen behandelingen onderling worden vergeleken zonder dat er sprake is van een niet-behandelde controlegroep. De keuze voor een bepaalde controlegroep wordt ingegeven door de vraag welk soort vergelijking beleidsmatig gezien het relevantst is en door de vraag of er überhaupt een "niet-behandelde" controlegroep kan worden gevonden. In tekstkader 10 hieronder staat een voorbeeld van een vergelijking tussen twee behandelingen zonder dat er sprake is van een "niet-behandelde" controlegroep, met de bedoeling om te beoordelen welke van de twee interventies wordt voorgezet. Verder moet worden opgemerkt dat onduidelijkheid kan ontstaan wanneer programma's onderling worden vergeleken zonder dat kan worden teruggegrepen op een niet-behandelde controlegroep (dit wordt in tekstkader 11 verder besproken).

*Een behandeling vergelijken met een situatie zonder behandeling of met een alternatief*

### Tekstkader 10 Beleidsvragen in verband met een opleidingsprogramma

Neem als voorbeeld een beleidsmaker die een nieuwe, uit het ESF te financieren opleidingsinterventie wil invoeren (we noemen dit interventie A). Veronderstel verder dat er in de lidstaat al een opleidingsprogramma (B) bestaat dat op dezelfde personen is gericht maar dat uit nationale middelen wordt gefinancierd. In dat geval luidt een mogelijke beleidsvraag: liggen de arbeidsparticipatie en het inkomensniveau van de deelnemers aan interventie A na deelname boven die van de deelnemers aan interventie B? En, in het verlengde daarvan: biedt interventie A een betere prijs-kwaliteitverhouding? Als het inkomen van de deelnemers aan interventie A hoger ligt, is de voor de hand liggende beleidsreactie om interventie B te staken ten gunste van interventie A, mocht de uitvoering daarvan ook kosteneffectief blijken.

### Tekstkader 11 De netto-effecten interpreteren

Het kan voorkomen dat in het onderzoek geen inkomensverschillen tussen de deelnemers aan interventie A en aan interventie B naar voren komen. De beleidsreactie op deze informatie laat zich niet altijd raden, bijvoorbeeld wanneer interventie B zeer effectief blijkt ten opzichte van situaties waarin mensen geen behandeling ondergingen. Dit zou betekenen dat beide interventies uiterst effectief zijn. Soms bestaan er evenwel geen aanwijzingen met betrekking tot de effectiviteit van interventie B ten opzichte van een situatie

zonder behandeling. Ook is het mogelijk dat interventie A en B beide ineffectief zijn, hoewel de ene interventie relatief gezien misschien effectiever lijkt dan de andere. Ook in situaties waarin bepaalde bevolkingsgroepen in meerdere interventies centraal staan, kan het nog altijd nuttig zijn om op zoek te gaan naar een geschikte groep niet-behandelde eenheden die als vergelijkingsmateriaal kan fungeren.

Merk op dat zonder niet-behandelde controlegroep de difference-in-differencesmethode niet kan worden gebruikt voor het vergelijken van meerdere behandelingen.

### **Welke praktische overwegingen moeten bij het samenstellen van een controlegroep een rol spelen?**

Naast overwegingen van analytische en beleidsmatige aard moet bij het samenstellen van controlegroepen ook rekening worden gehouden met praktische aspecten. Om al dan niet steekproefsgewijs eenheden (personen of ondernemingen) te selecteren die als controle-eenheid kunnen fungeren, moet naar een geschikt steekproefkader worden gezocht. Daarnaast moeten de steekproefkaders bestaan uit individuele eenheden die aan de analytische en beleidsmatige eisen beantwoorden. Wat hiervoor de beste methode is, verschilt per evaluatie en is afhankelijk van de specifieke omstandigheden waaronder de interventie wordt uitgetoetst.

Voor het bepalen van geschikte controlegroepen worden vaak twee gegevensbronnen gebruikt. Voor beide gegevensbronnen geldt dat de identiteit van de behandelingsgroep bekend moet zijn.

Bij het zoeken naar controle-eenheden kan gebruik worden gemaakt van bevolkingsregisters van uiteenlopende aard. Bij een actieve arbeidsmarktinterventie die bijvoorbeeld is gericht op mensen in de leeftijdsgroep 18-24 jaar met een werkloosheidsuitkering, kunnen de uitkeringsgegevens van deze mensen worden gebruikt om de doelpopulatie te bepalen. Als daarnaast de personen in de behandelingsgroep bekend zijn en met de uitkeringsgegevens kunnen worden gematcht, kunnen de 18-24-jarigen die niet worden behandeld — en die dus kandidaat zijn voor de controlegroep — worden opgespoord. Of neem een interventie die op kleine en middelgrote ondernemingen is gericht. Voor het bepalen van de doelpopulatie kunnen landelijke ondernemingsregisters (indien beschikbaar) worden gebruikt. Aan de hand van de informatie die beschikbaar is over de behandelde ondernemingen kunnen kandidaten voor de controlegroepen worden opgespoord.

*Bevolkingsregisters en fiscale gegevens over ondernemingen*

Als de interventie niet generiek wordt toegepast, kunnen de gegevens van de deelnemers worden gebruikt, bijvoorbeeld wanneer niet iedereen die zich voor een opleidingsprogramma heeft opgegeven, wordt toegelaten (een controlegroep met keuze/besef). Evenmin worden alle financieringsaanvragen van ondernemingen ingewilligd. De ondernemingen die niet voor een opleiding of voor financiering in aanmerking komen, kunnen in sommige gevallen als controle-eenheden dienen (zie eerder in dit gedeelte waar wordt ingegaan op het punt dat bij het kiezen van controlegroepen onder deze omstandigheden voorzichtigheid geboden is).

*Gegevens van de deelnemers*

### **2.3.4. Welke soorten gegevenskwesties moeten in het evaluatieschema aan de orde komen?**

**Welke soorten gegevens zijn vereist en hoe worden zij verzameld?**

Zoals gezien, vereisen CIE's doorgaans toegang tot aanzienlijke hoeveelheden microgegevens (in sommige gevallen kunnen gegroepeerde gegevens worden gebruikt, bijvoorbeeld regionale gegevens). Deze gegevens moeten worden verzameld, geordend en gedocumenteerd; van verschillende bronnen afkomstige gegevens moeten met elkaar worden verbonden op basis van gedeelde identificerende velden; zij moeten veilig worden opgeslagen en overgedragen tussen degenen die het CIE beheren en uitvoeren; en op basis van deze gegevensbronnen moeten analytische datasets worden samengesteld om de schatting van de effecten te vergemakkelijken.

*Het beheer van gegevensbronnen*

Bij de ontwikkeling van een evaluatieschema is het van belang rekening te houden met de volgende gegevensgerelateerde kwesties:

*Een basischecklist voor gegevensbeheer*

#### *Beschikbaarheid*

- Welke bronnen kunnen worden gebruikt om verschillende soorten gegroepeerde/microgegevens te verkrijgen?
- Zijn de vereiste individuele gegevens beschikbaar? Geldt dit ook voor gegevens van “bijzondere categorieën” (indien vereist)?
- Wanneer er een enquête moet worden gehouden, zijn de doelgroep en de controlegroep dan zodanig geïdentificeerd dat zij via enquête-interviews kunnen worden gevolgd — zijn hun contactgegevens beschikbaar en actueel?

#### *Samenhang*

- Is er sprake van één gegevensbron beschikbaar of moeten meerdere gegevensbronnen aan elkaar worden gekoppeld (bv. statistieken over werkloosheid, sociale uitkeringen, sociale zekerheid, bedrijfs-/vestigingsgegevens enz.)?
- Zijn de bronnen onderling samenhangend? Kunnen personen in de bronnen op coherente wijze worden geïdentificeerd?

#### *Toegankelijkheid*

- Is het mogelijk om toegang te krijgen tot nationale gegevensbronnen over individuele loopbanen, inkomen of sociale uitkeringen om ESF-deelnemers te vergelijken met een potentiële controlegroep?

#### *Overeenkomsten*

- Bestaan er tussen de beheerautoriteit en andere gegevenseigenaren specifieke overeenkomsten inzake de beschikbaarheid van gegevens? Welke juridische of organisatorische belemmeringen moeten worden besproken?
- Wie is verantwoordelijk voor het onderhandelen over toegang en het bereiken van overeenstemming over het gebruik van gegevens?

#### *Opslag*

- Waar worden de gegevens opgeslagen? Welke IT-systemen en -infrastructuur zijn nodig?
- Welke stappen worden genomen om ervoor te zorgen dat de gegevens veilig worden opgeslagen en dat de toegang is voorbehouden aan degenen die de gegevens nodig hebben voor onderzoek?
- Hoe worden gegevens geanonimiseerd? Kunnen personen door de tijd heen en via aan hen gekoppelde gegevensbronnen worden gevolgd?

*Naleving van gegevensbescherming*

- Voldoen eerdere oplossingen aan de basisregels van de AVG en de ESF+-verordening? Zijn incidentele belemmeringen voor de toegankelijkheid van gegevens gerechtvaardigd door het gebruik van verschillende rechtsgrondslagen overeenkomstig artikel 6, punten a), c) en e), AVG en artikel 17 van de ESF+-verordening en het mogelijke gebruik van geanonimiseerde gegevens?

**Hoe worden de gegevens verwerkt?**

In veel gevallen zijn voor CIE's microgegevens vereist, d.w.z. gegevens die waarnemingen bevatten met betrekking tot afzonderlijke eenheden (normaal gesproken personen of ondernemingen) in zowel de behandelings- als de controlegroep (soms kunnen gegroepeerde gegevens worden gebruikt, bv. gegevens op regionaal of ODA-niveau). Wij hebben een onderscheid gemaakt tussen de drie belangrijkste soorten vereiste gegevens: a) gegevens over de behandelings- en de controlegroep, b) gegevens over de resultaten, en c) contextuele gegevens (gegevens die worden gebruikt om belangrijke potentiële verschillen tussen de behandelings- en de controlegroep te controleren). Deze gegevens kunnen afkomstig zijn van dezelfde bron of van verschillende bronnen. De bronnen moeten zodanig zijn gestructureerd dat zij analytische datasets (of analytische steekproeven) vormen die kunnen worden gebruikt om effecten te schatten. Deze structurering houdt in veel gevallen in dat gegevens van afzonderlijke personen of ondernemingen uit verschillende bronnen worden gekoppeld. Een dergelijke koppeling vereist ofwel afzonderlijke identificatiemiddelen (bijvoorbeeld individuele socialezekerheidsnummers), waardoor het dossier van een persoon (bijvoorbeeld in de belastinggegevens) kan worden afgestemd op de deelnamegegevens, ofwel voldoende gegevens om dossiers uit verschillende bronnen te koppelen (bijvoorbeeld naam en geboortedatum moeten in die bronnen beschikbaar zijn). Het is belangrijk om na te gaan welke gegevensbronnen voor het geplande CIE zullen worden gebruikt, maar ook om stil te staan bij de vraag of het mogelijk is om gegevens uit verschillende bronnen te koppelen.

*Het koppelen van microgegevens uit verschillende bronnen*

**2.3.5. Wat zijn de belangrijkste beperkingen bij het analyseren van gegevens en resultaten?**

Zoals hierboven besproken, worden de effecten in de CIE's gewoonlijk bepaald door de resultaten in de behandelingsgroep te vergelijken met die in de controlegroep, en deze fundamentele vergelijking maakt in wezen deel uit van de CIE-benadering. Het verschil tussen de behandelings- en de controlegroep wordt het netto-effect van de interventie genoemd. De precieze wijze waarop de netto-effecten worden geschat, zal echter afhangen van de gekozen onderzoeksopzet.

Bij de planning van een CIE is het van belang om na te gaan of de interventie groot genoeg is en vermoedelijk effecten zal sorteren die statistisch aan het licht kunnen komen. De CIE-methoden zijn namelijk gebaseerd op het beginsel dat de doelgroep en de controlegroep steekproeven zijn van de gehele potentieel behandelde en controlepopulaties en dat, om betekenisvolle resultaten te behalen, het aantal personen in de twee groepen statistisch significant moet zijn.

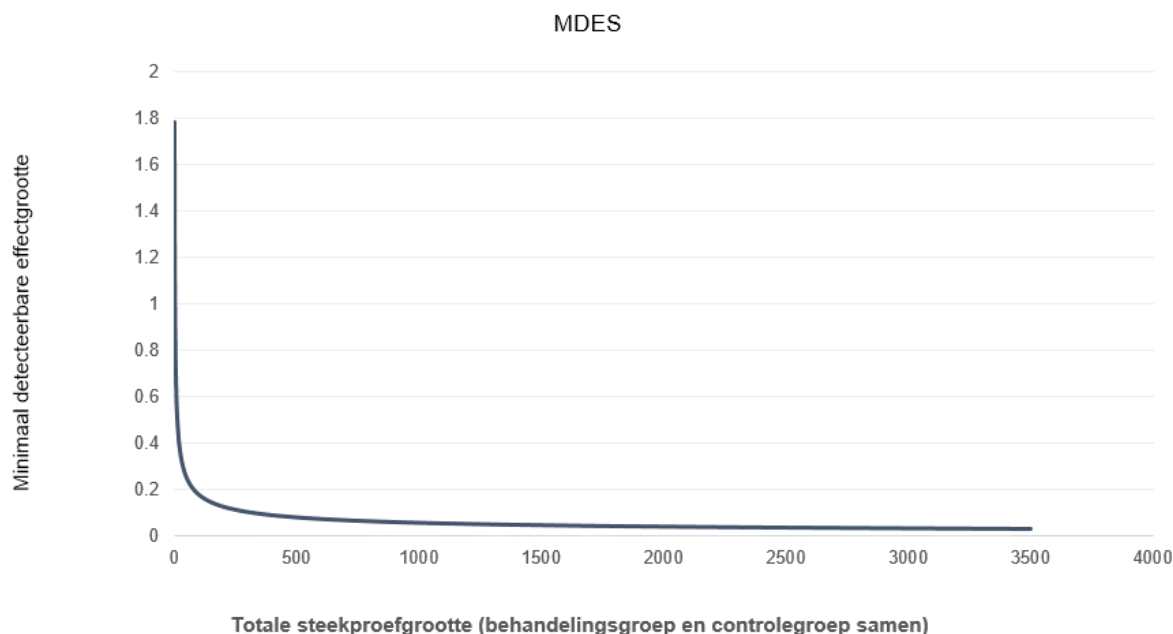


Bij de beoordeling of een steekproef groot genoeg is voor analyse, is “minimaal detecteerbaar effect” een nuttig begrip<sup>65</sup>. Eenvoudig gezegd is een minimaal detecteerbaar effect het kleinste werkelijke effect dat een steekproefgrootte kan detecteren bij standaardniveaus van statistisch vertrouwen en statistisch vermogen<sup>66</sup>.

*Beoordeling van steekproefgrootte en effectgrootte*

Figuur 4 laat zien hoe de grootte van het minimaal detecteerbare effect varieert naargelang van de totale steekproefgrootte (totaal aantal steekproeven in de behandelingsgroep en de controlegroep) bij een statistisch vertrouwen van 95 % en een statistisch vermogen van 80 %. Van links naar rechts neemt de grootte van het minimaal detecteerbare effect snel af naarmate de steekproefgrootte de 500 eenheden (250 behandelingseenheden en 250 controle-eenheden) nadert. Met andere woorden, naarmate de totale steekproefgrootte toeneemt, is de CIE-opzet nauwkeuriger en in staat om kleinere effecten te detecteren.

**Figuur 4 Minimaal detecteerbare effectgrootten (MDES) bij verschillende steekproefgrootten**



Opmerking: 95 % statistische significantie en 80 % statistisch vermogen. In de figuur wordt een gerandomiseerde opzet getoond.

Bij de planning van een CIE is het nuttig een schatting te maken van de vermoedelijke grootte van de steekproef om de behandelde groep en de controlegroep te bepalen; deze schatting is gebaseerd op prognoses van het aantal eenheden dat zal worden behandeld, de methode die waarschijnlijk in

*De waarschijnlijke grootte van de steekproef*

<sup>65</sup> Bloom H. S., 1995 verstrekt praktische richtsnoeren voor het berekenen van minimale detecteerbare effecten voor experimentele opzetten. In geval van quasi-experimentele benaderingen zullen dergelijke berekeningen moeten worden aangepast. Over het algemeen vereisen quasi-experimentele methoden een grotere steekproefgrootte dan die welke nodig is voor een experimentele opzet.

<sup>66</sup> Het niveau van statistisch vertrouwen is een statistische maatstaf voor de betrouwbaarheid van de schattingsprocedure, terwijl het niveau van statistisch vermogen de waarschijnlijkheid is dat een test een werkelijk effect aantoonst. Deze twee parameters hangen ook af van de steekproefgrootte en kunnen omgekeerd worden gebruikt om een steekproefgrootte te berekenen met het oog op geschikte significantieniveaus. Over het algemeen is het statistisch vertrouwen 95 % en het statistisch vermogen 80 %.

het CIE zal worden gebruikt en de omvang van de overeenkomstige controlegroepen op basis van beschikbare administratieve gegevens of een passende enquête. Deze informatie kan daarna op basis van bepaalde aannames worden gebruikt om na te gaan of de resulterende minimale detecteerbare effecten voldoende significant zijn en of de uitvoering van het CIE een redelijke keuze is. Een belangrijk aspect hierbij is of de betrokken interventie waarschijnlijk effecten zal voortbrengen die even groot zijn als de geschatte minimaal detecteerbare effecten; In dit verband kan een studie van de bestaande literatuur en andere vergelijkbare evaluaties nuttig zijn.

Er wordt ook op gewezen dat met CIE's vaak wordt beoogd de effecten op verschillende bij de interventie betrokken subpopulaties te onderzoeken (mannelijk/vrouwelijk, jong/oud, minder/beter opgeleid enz.). Deze onderzoeken verminderen onvermijdelijk de omvang van de behandelde groep en de controlegroep voor elke subpopulatie (mannelijk, vrouwelijk, jong enz.) met als gevolg dat de statistische significantie afneemt. Wanneer de studie van de effecten in de subpopulaties een belangrijk onderdeel van de evaluatie is, moet elke subpopulatie beschikken over een voldoende aantal eenheden in de behandelings- en de controlegroep.

Voor de vaststelling van de steekproefgrootte zijn technische vaardigheden vereist waarover de beheerautoriteit niet altijd beschikt. In dat geval moet de beoordelaar de steekproefgrootte berekenen en overeenstemming bereiken met de beheerautoriteit over de haalbaarheid en de reikwijdte van het CIE. Het is niet mogelijk de drempelwaarde voor de steekproefgrootte vast te stellen waarboven het altijd mogelijk is een CIE uit te voeren, omdat deze drempel kan verschillen afhankelijk van de gebruikte methoden, de verwachte onderzoeken van subpopulaties, de aanvaarde niveaus van statistische significantie, en andere elementen. In veel evaluaties bleek echter een aantal van 2 000 personen voor de behandelingsgroep en de controlegroep samen voldoende te zijn voor een redelijk nauwkeurige analyse (met inbegrip van de belangrijkste subgroepen, zoals mannen/vrouwen enz.). Dit cijfer kan niet worden gebruikt als een wetenschappelijke "drempel", maar wel als hulpmiddel bij het onderzoeken van de mogelijke uitvoering van een CIE.

De steekproefgrootte is van invloed op de robuustheid van de CIE-resultaten. Enkele voorbeelden van onzekerheden bij de interpretatie van de resultaten en het verband ervan met de steekproefgrootte worden in het onderstaande tekstkader toegelicht.

### **Tekstkader 12 Onzekerheid bij de interpretatie van de resultaten**

Onder de voorbeelden van evaluaties van uit het ESF gefinancierde interventies, leverde de evaluatie van verschillende maatregelen voor langdurig werklozen in de regio Marche in Italië geen statistisch significante resultaten op bij een van de vier onderzochte maatregelen, "Vouchers voor opleiding". Dit is waarschijnlijk te wijten aan het feit dat de steekproef van de onderzochte behandelde personen klein was in vergelijking met het grote aantal daadwerkelijke begunstigen. Iets soortgelijks deed zich voor bij de evaluatie van maatregelen ter bevordering van het beroepsonderwijs in Polen, in de provincie Podlaskie. Naast het probleem van de statistische significantie blijkt uit dit voorbeeld dat, bij de analyse van de effecten, het gebruik van te kleine steekproeven van behandelingsgroepen kan leiden tot problemen op het gebied van de generalisatie van de bevindingen.

De evaluatie van taalcursussen voor immigranten in Duitsland toont positieve resultaten, in tegenstelling tot een aantal eerdere studies waarin de effecten van taalopleidingsprogramma's voor immigranten werden beoordeeld. Volgens de auteur is een mogelijke verklaring voor dit verschil het feit dat het Duitse programma in kwestie ook een module werkervaring omvatte. Deze hypothese kon door de auteur aan de

hand van de beschikbare gegevens niet worden getoetst, omdat het onmogelijk was een onderscheid te maken tussen de onderdelen taalopleiding en werkervaring van de interventie. Dit kan interessant zijn voor toekomstig onderzoek. Uit het voorbeeld blijkt dat contrafeitelijke onderzoeken soms moeten worden herhaald met verbeterde en uitgebreidere gegevens of dat zij moeten worden aangevuld met andere evaluatiemethoden om uit te leggen welke mechanismen bepalend zijn voor de resultaten van een interventie.

### 2.3.6. Een checklist om de voorbereiding en haalbaarheid van het CIE te verifiëren

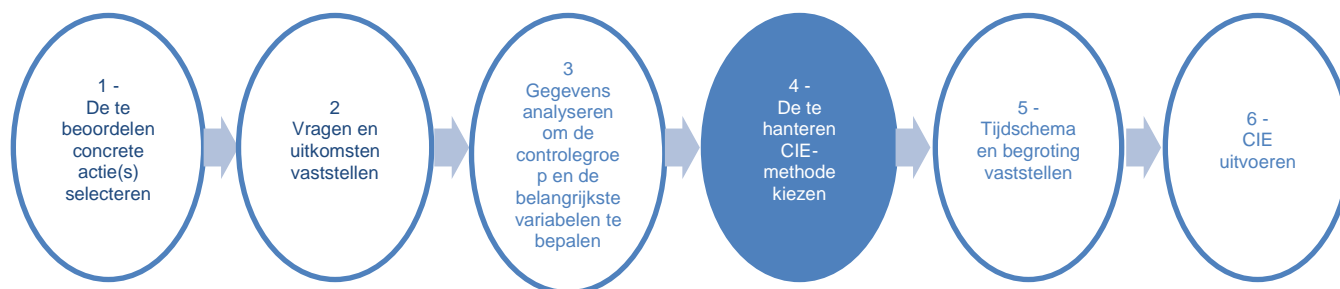
Nu is het nuttig om in een eenvoudige checklist de belangrijkste factoren samen te vatten die een beheerautoriteit bij de voorbereiding van een CIE in aanmerking moet nemen en verifiëren:

*Een lijst van belangrijke factoren waarmee rekening moet worden gehouden*

- Is de voor evaluatie gekozen interventie geschikt voor een CIE?
  - o Is er sprake van een discrete, onderscheidende en relatief homogene interventie?
  - o Wijst de interventielogica op een overtuigend causaal mechanisme van de uitkomsten die door een CIE moeten worden onderzocht?
  - o Kan een controlegroep gemakkelijk worden vastgesteld overeenkomstig de regels van de interventie?
- Zijn de effecten op de deelnemers kwantitatief meetbaar?
  - o Zijn de gemeten uitkomsten van de interventie duidelijk en in overeenstemming met de interventielogica?
  - o Is er voldoende tijd verstreken sinds het einde van de behandeling om de uitkomsten voor de deelnemers te detecteren?
- Is er in verband met de evaluatievragen een meting van de netto-effecten van de interventie vereist?
- Zijn er geschikte gegevens (gegevens over de behandelingsgroep en de controlegroep, uitkomstgegevens, contextgegevens) beschikbaar of kunnen zij beschikbaar worden gesteld?
  - o Voldoen de beschikbare gegevens aan de AVG? Zo niet, kunnen dan de nodige overeenkomsten en technische oplossingen tussen administratieve instanties worden geactiveerd om de gegevens conform te maken?
  - o Is het mogelijk om met de beschikbare gegevens een geschikte controlegroep te bepalen?
  - o Is de steekproef groot genoeg om binnen de CIE-resultaten de vereiste statistische significantie te bereiken?

Een positief antwoord op al deze vragen houdt in dat het mogelijk is een CIE op te zetten en uit te voeren. Een negatief antwoord op enkele van deze vragen sluit de uitvoering van een CIE niet uit, maar vereist aanvullende controles van de haalbaarheidsvoorwaarden en verbeteringen in de beschikbaarheid van gegevens of andere belangrijke factoren.

## 2.4. Toe te passen CIE-methode



Het kiezen van de in het CIE te gebruiken methode is een cruciale stap in de voorbereiding van de evaluatie, aangezien de gekozen methode zowel de kwaliteit van de bevindingen als de algehele kwaliteit van de evaluatie beïnvloedt. Algemeen gezegd is geen enkele methode beter dan een andere, maar kan een bepaalde methode geschikter zijn om de beschikbare gegevens te gebruiken, terwijl een andere methode geschikter is voor het onderzoeken van een bepaald type interventie. De vaststelling van de meest geschikte methode voor een bepaalde evaluatie is een cruciale stap in de opzet.

*Enkele basisvoorwaarden voor het bepalen van de meest geschikte methode*

De technische aspecten en de intrinsieke kenmerken van elke methode worden uitvoerig behandeld in **hoofdstuk 3**, waarbij het hier nuttig is kort stil te staan bij de praktische gevolgen van de keuze.

Ten eerste moet een gerandomiseerde aanpak — volgens welke personen willekeurig aan de behandelingsgroep of de controlegroep worden toegewezen — vóór het begin van de interventie worden vastgesteld. Dit veronderstelt zowel de voltooiing van de evaluatieopzet als een organisatie die in staat is de taak uit te voeren. Bij ESF-interventies wordt deze methode zelden gehanteerd, maar zoals wij zullen zien, is zij onder bepaalde voorwaarden mogelijk en zeer doeltreffend vanuit methodologisch oogpunt.

Ten tweede kan de keuze van de CIE-methode de capaciteiten van de beheerautoriteit te buiten gaan, omdat hiervoor geavanceerde technische vaardigheden en specifieke ervaring nodig zijn. In dat geval kiest de beoordelaar de meest geschikte methode, maar natuurlijk pas nadat hij of zij is benoemd. Dit kan het bijkomende voordeel hebben dat het als criterium kan worden gebruikt om de bekwaamheid van beoordelaren te beoordelen tijdens de selectieprocedure van deze beoordelaren. In ieder geval moet de beheerautoriteit vertrouwen hebben in de uitvoering van een CIE nadat zij alle of de meeste vragen uit de vorige checklist positief heeft beantwoord en — intern of extern — de nodige vaardigheden heeft verworven om de keuze van de beoordelaar goed te keuren.

Ten derde kunnen gerandomiseerde methoden verschillen van quasi-experimentele methoden wat betreft duur en kosten, en moet de beoordelaar vanaf het begin van de interventie worden betrokken bij gerandomiseerde onderzoeken met een controlegroep. Quasi-experimentele methoden verschillen niet sterk van elkaar; de beoordelaar kan zijn of haar werkzaamheden starten nadat de interventie is begonnen (maar tijdig genoeg om de nodige gegevens te organiseren en voor te bereiden) en de kwaliteit van de beschikbare gegevens zou het verschil kunnen maken wat betreft tijd en kosten.

Tot slot maken het type interventie en de beschikbaarheid van gegevens soms het gebruik van meer dan één quasi-experimentele methode mogelijk. In dit geval kan het nuttig zijn om meer methoden toe te passen omdat de bevestiging van dezelfde effecten met verschillende methoden de robuustheid van de bevindingen verbetert.

## 2.5. Tijdschema en begroting



### 2.5.1. Welke middelen zijn beschikbaar?

Een belangrijk aandachtspunt bij de uitwerking van een CIE-evaluatieschema is de beschikbaarheid van middelen. Dit kan een breed scala van overwegingen omvatten, die hier in drie categorieën zijn onderverdeeld: a) deskundigen; b) tijd; en c) financiële middelen.

#### Welke externe deskundigen en interne personeelsleden zijn er nodig voor een CIE?

In de meeste gevallen zal een effectonderzoek worden uitbesteed aan een externe leverancier. De opdracht moet echter binnen de beheerautoriteit worden beheerd door personeel met kennis van de CIE-methoden. Deze kennis is nodig om de kwaliteit te waarborgen en doeltreffend contact te onderhouden met externe deskundigen. Binnen de beheerautoriteit kunnen ook andere vormen van deskundigheid vereist zijn, zoals statistische vaardigheden en deskundigheid op het gebied van gegevensverzameling en -beheer. Het is belangrijk om vooraf na te gaan of de beheerautoriteit over voldoende gekwalificeerd en opgeleid personeel beschikt en of dit personeel in staat is de evaluatie te ondersteunen.

*Intern personeel*

Voor de opdracht van een doeltreffend CIE zijn contractanten vereist die over de nodige vaardigheden en ervaring beschikken om dergelijke evaluaties uit te voeren. Bovendien moeten geschikte contractanten de beleids- en administratieve context in de lidstaten begrijpen, vertrouwd zijn met potentiële gegevensbronnen en de juiste talen beheersen. Het is de moeite waard na te gaan wat er nodig is om binnen een lidstaat een CIE-leveranciersbasis te ontwikkelen (zie hoofdstuk 4 voor een nadere behandeling van dit onderwerp).

*Extern personeel*

Doeltreffende CIE's vereisen de medewerking van de beheerders van de onderzochte programma's of interventies. Zo is bijvoorbeeld toegang vereist tot de registers die door de beheerders van de interventie worden bijgehouden. Deze registers verschaffen informatie over personen of ondernemingen die aan een interventie hebben deelgenomen.

*Personeel dat de programma's / interventies beheert*

Beheerders van programma's/interventies kunnen advies en richtsnoeren geven met betrekking tot dit soort gegevens. Zij kunnen ook worden verplicht

om registers bij te houden naast wat zij zouden moeten doen wanneer er geen effectonderzoek plaatsvindt.

Om het probleem van het verzamelen van gegevens uit verschillende bronnen op te lossen, zullen degenen die een CIE plannen contact moeten onderhouden met het personeel dat zich bezighoudt met officiële gegevensbronnen (bv. werkloosheidsregister, socialezekerheidsgegevens, bureaus voor de statistiek enz.). Hierdoor zal het mogelijk zijn de gegevensverstrekking ruimschoots van tevoren te plannen.

*Statistische deskundigheid*

### **Welke factoren zijn van belang voor het opstellen van een tijdschema voor een CIE?**

Voor de uitvoering van een CIE zijn bijdragen van verschillende personeelsleden nodig; daarnaast worden dergelijke evaluaties over lange perioden uitgevoerd. Een evaluatieschema moet een ontwerptijdschema omvatten met cruciale projectmijlpalen, die gelden voor degenen die betrokken zijn bij de interventie zelf alsook voor degenen die betrokken zijn bij de evaluatie. Het ontwerptijdschema moet betrekking hebben op zowel de evaluatie- als de interventie-activiteiten, alsook op de belangrijkste beleidsmijlpalen.

*CIE's vergen tijd*

Het ontwikkelen van een zinvol en realistisch tijdschema voor een CIE is een moeilijke evenwichtsoefening. Enerzijds moet de beheerautoriteit (of de intermediaire instantie) die de evaluatie plant, rekening houden met de belangrijkste data waarop beslissingen moeten worden genomen die gebaseerd zijn op de bevindingen van de evaluatie. Anderzijds zullen er onvermijdelijk beperkingen zijn die de timing van de verslagen beïnvloeden. Sommige resultaten zullen pas na jaren naar boven komen en dit zal zoveel mogelijk tot uiting moeten komen in de tijdschema's voor het verzamelen, analyseren en rapporteren van gegevens (zie punt 2.5.2 hieronder).

*Het plannen van een tijdspanne waarin effecten kunnen ontstaan*

De evaluatie moet in een vroeg stadium van de programmeringsperiode worden uitgevoerd, zodat eventuele veranderingen kunnen worden doorgevoerd en in de resterende tijd gebruik kan worden gemaakt van de opgedane ervaringen en geleerde lessen. In bepaalde omstandigheden kunnen dezelfde of soortgelijke interventies in opeenvolgende programmeringsperioden worden ondersteund. Resultaten van CIE's die gericht zijn op interventies uit voorgaande programmeringsperioden kunnen uitermate nuttig zijn als informatie voor de uitvoering en opzet in latere programmeringsperioden.

*Focus op specifieke punten in de tijd*

Ook is het van belang om na te gaan hoe de timing van een contrafeitelijk effectonderzoek in verband kan worden gebracht met de timing van andere onderdelen van de evaluatie. Waarschijnlijk moet voorafgaand aan een CIE een theoriegebaseerde evaluatie worden uitgevoerd. Voor innovatieve interventies (bv. de ESF-interventies 2014-2020 ter ondersteuning van kinderen in Tsjechië, of de diensten ter bevordering van de sociale inclusie van personen die de nieuwe inkomenssteun in Italië ontvangen, of de verschillende maatregelen ter ondersteuning van werknemers en studenten tijdens de COVID-19-lockdown in veel lidstaten) kan het voorkomen dat eerder verslag moet worden uitgebracht over belangrijke elementen van een procesevaluatie om verbeteringen in het ontwerp van een CIE mogelijk te maken, of juist later om een gedetailleerd onderzoek naar oorzaken en effecten te vergemakkelijken. Wanneer een CIE van een volwassen lopende

*De volgorde van verschillende soorten evaluaties bepalen*

interventie wordt uitgevoerd, zou het relevanter zijn om de procesevaluatie naast het effectonderzoek uit te voeren.

Een tijdschema zal ook afhankelijk zijn van de beschikbaarheid van gegevens. Het bijwerken van gegevensbronnen kan veel tijd in beslag nemen, zoals vaak het geval is bij belastinggegevens. Ook juridische en institutionele belemmeringen bij het verkrijgen van de benodigde gegevens kunnen tijdrovend en duur zijn. Bovendien vergt het gebruik van gegevens uit uiteenlopende bronnen, het waarborgen van de compatibiliteit, het controleren van de kwaliteit en het vormgeven van deze gegevens in een formaat dat kan worden gebruikt om de effecten te schatten, veel tijd en moeite.

*Het verzamelen van gegevens kan tijdrovend zijn*

### **Hoe kunnen de kosten worden geraamd?**

Het is belangrijk een indicatieve begroting op te stellen van het bedrag dat de beheerautoriteit in staat en bereid is aan een CIE te besteden. De begroting bestaat uit twee onderdelen: interne en externe kosten. De eerste hebben betrekking op de inspanningen van het interne personeel dat de evaluatie volgt; binnen overheidsdiensten zijn deze kosten over het algemeen niet in financiële termen uitgedrukt, maar moeten zij in aanmerking worden genomen om een passende controle en follow-up van de evaluatie te waarborgen. De kosten voor het inschakelen van externe deskundigen voor de uitvoering van een CIE moeten zorgvuldig worden geraamd om te zorgen voor een hoogwaardige evaluatie tegen een redelijke prijs. De nadruk ligt hier op externe kosten.

Er moet een onderscheid worden gemaakt tussen evaluaties van routinematige interventies, waarbij de uitgaven doorgaans lager zijn, en innovatieve of proefprojecten waarvoor het verzamelen van een relatief grote hoeveelheid gegevens, het gebruik van nieuwe gegevensbronnen en de betrokkenheid van nieuwe belanghebbenden hogere uitgaven kunnen rechtvaardigen. Dit is echter geen vaste regel en moet worden gezien in de context van de evaluatievragen, waarvan de complexiteit en het aantal minder of meer financiële middelen vereisen.

*Een mogelijke methode voor kostenraming*

Ook de keuze van de evaluatiemethode maakt een verschil. Zoals eerder opgemerkt, vereist een gerandomiseerde methode de aanwezigheid van de beoordelaar om informatie te verzamelen en te controleren tijdens de duur van de interventie. Anderzijds wordt de inspanning op het gebied van quasi-experimentele methoden grotendeels bepaald door het aantal gegevensbronnen en de kwaliteit en beschikbaarheid hiervan.

In een leidraad van de Commissie<sup>67</sup> wordt een indicatie gegeven van het bedrag dat nodig is voor de evaluatie van een programma en wordt gesteld dat grootschalige en routinematige programma's niet meer dan 1 % van hun programmabegroting aan evaluaties mogen besteden, terwijl voor innovatieve of proefinitiatieven tot 10 % mag worden besteed. In deze leidraad wordt echter niet expliciet ingegaan op de vereiste middelen van de CIE's en wij kunnen ervan uitgaan dat deze percentages het plafond vormen voor evaluatiekosten, met inbegrip van een of meer CIE's.

Zoals men zich kan voorstellen, is het onmogelijk om indicatieve kosten van een CIE te geven voor een dermate groot aantal verschillende interventies en operationele omstandigheden als die van de ESF+-programma's in

<sup>67</sup> Zie [Europese Commissie, 2013](#)

27 lidstaten. In dit geval is het nuttiger een methode voor te stellen voor het ramen van een redelijke begroting voor CIE's in verschillende contexten. In de volgende tabel is een simpele aanpak weergegeven: de rijen geven een overzicht van de belangrijkste uit te voeren taken, terwijl in de kolommen de belangrijkste kosten worden vermeld.

De duur en complexiteit van elke activiteit bepalen de vereiste inspanning uitgedrukt in werkdagen van het evaluatieteam en dus de belangrijkste kosten ervan. De kosten van gegevensverzameling door middel van kwantitatieve enquêtes onder de deelnemers en de leden van de controlegroep zijn aanzienlijk. Wanneer een CIE is gebaseerd op bestaande administratieve gegevensbronnen en deze relatief gemakkelijk toegankelijk zijn, zullen de totale kosten lager zijn. Administratieve gegevens vereisen echter vaak een essentiële "voorbereidende" activiteit (correctie van ontbrekende of onjuiste gegevens, aanpassing van het formaat van de databank, controle van de administratieve voorschriften inzake gegevens enz.), waardoor de kosten kunnen oplopen.



**Tabel 3 Structuur van de belangrijkste kosten van een CIE**

Voornaamste activiteiten	Evaluatieteam			Andere kosten (Apparatuur, materialen, benodigdheden, reizen enz.)	Toelichting
	Programma beheerder	Senior deskundigen	Junior deskundigen		
Planning en coördinatie					Het planningsproces omvat de haalbaarheid van het CIE, de organisatie ervan en de bepaling van de methodologie in overleg met de beheerautoriteit. In deze fase moet voldoende tijd worden besteed aan de analyse van mogelijke lacunes in de gegevens. De coördinatie bestrijkt de gehele duur van het CIE en omvat de organisatie van de werkzaamheden en de interactie met de beheerautoriteit en andere betrokken belanghebbenden om de evaluatievragen af te ronden.
Literatuuronderzoek					Het literatuuronderzoek dient ter ondersteuning van het schetsen van een beeld van de interventielogica, het vaststellen van de te gebruiken variabelen en het begrip van de context. Het maakt het ook mogelijk de evaluatievragen te verfijnen, de resultaten uit het verleden te gebruiken en hypothesen te formuleren.
Verzameling en voorbereiding van gegevens					De voor deze activiteiten vereiste inspanningen variëren sterk, afhankelijk van de methode(n) van gegevensverzameling. Het verzamelen van originele gegevens via enquêtes kan tijdrovend en duur zijn. Voor andere methoden kan het nodig zijn te investeren in technologie (software of hardware) of overeenkomsten te sluiten met gegeveuseigenaren. De kosten kunnen worden verminderd door de gegeveuseigenaren te betrekken bij de voorbereiding van de gegevens.
Gegevensanalyse					Voor gegevensanalyse zijn geavanceerde vaardigheden en ervaring vereist. De tijd die nodig is voor deze activiteit wordt beïnvloed door de methode, het aantal analyses en de kwaliteit van de gegevens.
Opstelling verslag(en)					De inspanningen voor deze activiteit variëren afhankelijk van het aantal en type verslagen en andere communicatiemiddelen. De kosten kunnen naast de voorbereidingstijd het printen en grafisch ontwerp omvatten.
Follow-upvergaderingen					Follow-upvergaderingen en andere verspreidingsactiviteiten zijn een belangrijke stap in de verspreiding van de bevindingen. Onder "diverse kosten" kunnen onder meer de huur van de ruimte en van voedsel vallen.
<b>Totaal</b>					

Bij gebrek aan andere referenties kunnen de personeelskosten worden berekend op grond van de ESF+-kosten voor opleidings- of arbeidsbemiddelingsdiensten die in elk land zijn vastgesteld voor senior en junior opleiders of deskundigen. Deze vergoedingen zijn over het algemeen vergelijkbaar met die van een junior en senior onderzoeker en zijn, ook al hoeven zij niet noodzakelijkerwijs gelijkwaardig te zijn, een nuttige parameter die in aanmerking moet worden genomen. Een eenvoudige marktanalyse en een aantal interviews met enkele onderzoekers kunnen helpen om de verwachte kosten van het CIE te specificeren.

### **2.5.2. Wanneer moet de interventie worden geëvalueerd?**

Het is van cruciaal belang vast te stellen welk moment in de loop van een interventie het meest geschikt is om een effectonderzoek uit te voeren, alsook de kritieke aspecten met betrekking tot het moment waarop de resultaten moeten worden gemeten en de effecten moeten worden geschat.

#### **Wanneer moeten nieuwe en lopende interventies worden geëvalueerd?**

De discussie over wanneer het in de loop van een interventie passend is een contrafeitelijk effectonderzoek uit te voeren, zal afhangen van de vraag of het om een nieuwe of een volwassen, lopende regeling gaat. Een nieuwe interventie heeft meer tijd nodig om volwassen te worden en een stabiele staat te bereiken. Het uitvoeren van een CIE voordat dit plaatsvindt is voorbarig en kan misleidend bewijs opleveren. In het geval van nieuwe interventies is een eerste procesevaluatie die vóór een CIE wordt uitgevoerd, vaak nuttig om aanloopproblemen op te sporen en oplossingen aan te wijzen.

*De timing van nieuwe en lopende interventies verschilt*

Bij het bepalen van de optimale timing van een CIE voor een nieuwe interventie moet rekening worden gehouden met een reeks andere factoren, waaronder: maatregelen om ervoor te zorgen dat er geschikte gegevensbronnen beschikbaar zijn, de oprichting van een intern projectteam bestaande uit naar behoren opgeleid personeel en de aanstelling van een externe contractant. Bovendien zullen de behoeften van het besluitvormingsproces waarop de evaluatie uiteindelijk is gericht, een kritieke beperking vormen.

*Kwesties met betrekking tot de middelen*

De timing van een effectonderzoek voor een lopende interventie zal voornamelijk zijn gebaseerd op praktische en beleidsgerelateerde vereisten. De interventie zou al concreet gestalte moeten hebben gekregen en een mate van volwassenheid hebben bereikt, zodat een CIE passend zou zijn. Een ander punt dat in aanmerking moet worden genomen, is de aanwezigheid van andere hervormingen die gelijktijdig met de te evalueren interventie plaatsvinden. De effecten van deze hervormingen kunnen van invloed zijn op de impact van de betreffende interventie. Beleidsmakers zullen moeten nagaan of de eventuele andere hervormingen op het gebied van beleid relevant zijn voor beleidsbeslissingen die gebaseerd zijn op de resultaten van het CIE in kwestie.

ESF-evaluaties zijn doorgaans gericht op één programmeringsperiode. Met name in het geval van stabiele interventies die al in de voorgaande periode deel uitmaakten van het ESF-programma, kan het echter nuttig

*Meerdere programmeringsperioden*

zijn om een retrospectieve evaluatie van de voorgaande periode te combineren met een lopende evaluatie van de huidige periode om een langere looptijd van een interventie te bestrijken.

### **Wanneer moeten de resultaten worden gemeten en de effecten worden berekend**

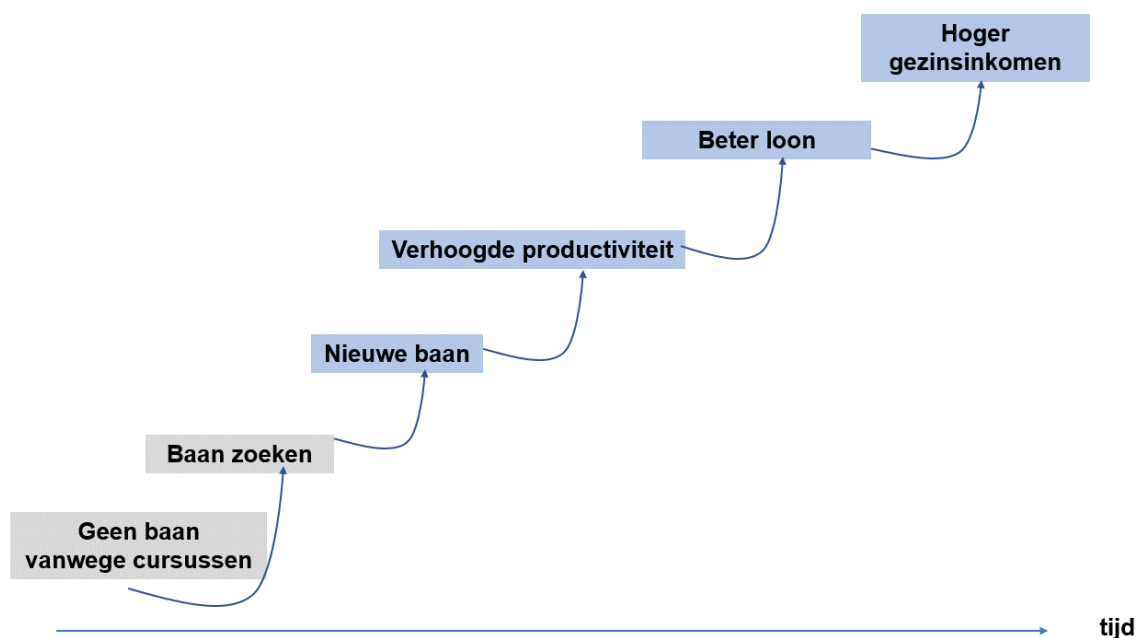
De tweede belangrijke kwestie in verband met de timing van een evaluatie is wanneer de effecten moeten worden gemeten en geschat of, meer in het bijzonder, wanneer wordt verwacht dat effecten zullen optreden na een interventie.

Wanneer een op werklozen gerichte opleidingsinterventie wordt onderzocht, rijst de vraag in welke periode een hogere arbeidsparticipatie kan worden verwacht. Het is een welbekend kenmerk van opleidingsprogramma's dat zij op de korte termijn de arbeidsparticipatie van de deelnemers verminderen. Dit is te wijten aan het zogenoemde "lock-in-effect". Interventies op het gebied van opleiding leiden er vaak toe dat werkloze stagiairs het zoeken naar werk onderbreken vanwege hun deelname aan cursussen. Personen uit de controlegroep daarentegen zetten zich ervoor in om een baan te vinden. Wanneer de effecten te vroeg worden berekend, kunnen ze negatief of ondergewaardeerd zijn. Bij de planning van een CIE is het van belang realistisch te zijn over de timing van de effecten en over het moment waarop zij waarschijnlijk kunnen worden gemeten. Figuur 5 bevat een vereenvoudigd model van een opeenvolging van effecten.

Bij de beoordeling van de vraag wanneer de resultaten het best kunnen worden gemeten en de effecten kunnen worden ingeschat, moet rekening worden gehouden met de termijnen waarbinnen de beleidsmakers informatie nodig hebben. In het geval van interventies die gericht zijn op het verbeteren van de inzetbaarheid op de lange termijn, kan het vanuit analytisch oogpunt zinvol zijn om de deelnemers te volgen gedurende twee jaar of langer nadat zij een behandeling hebben ondergaan, om na te gaan of hun inkomen en arbeidsparticipatie hoger zijn dan die van een vergelijkbare groep personen die die behandeling niet hebben ondergaan. Programmabeheerders hebben daarentegen de bevindingen vaak snel nodig, waardoor er een compromis moet worden bereikt tussen wat een redelijke follow-upperiode is vanuit het oogpunt van de interventie en de behoefte van de besluitvormers aan tijdige empirische gegevens.

*Wat is redelijk en wat is haalbaar*

**Figuur 5 Vereenvoudigd tijdschema voor de resultaten van een opleidingsprogramma**



Wanneer resultaatmetingen worden verkregen uit administratieve bronnen (bv. socialezekerheidsdossiers met gegevens over werk en inkomen), dan is het praktisch om de resultaten over een langere periode herhaaldelijk te volgen en de effecten te schatten (zelfs maandelijks). In dit geval bestaat het risico dat de aard van de bevindingen in de loop van de tijd kan veranderen. Indien de verzameling van primaire gegevens in de vorm van steekproefenquêtes is vereist voor het meten van resultaten, zou het op gezette tijden schatten van de effecten zeer duur worden, tenzij retrospectieve gegevens over de resultaten op een haalbare manier kunnen worden verzameld. De kosten van het verkrijgen van gegevens uit meerdere administratieve systemen en het creëren van één analytische dataset mogen echter niet worden onderschat.

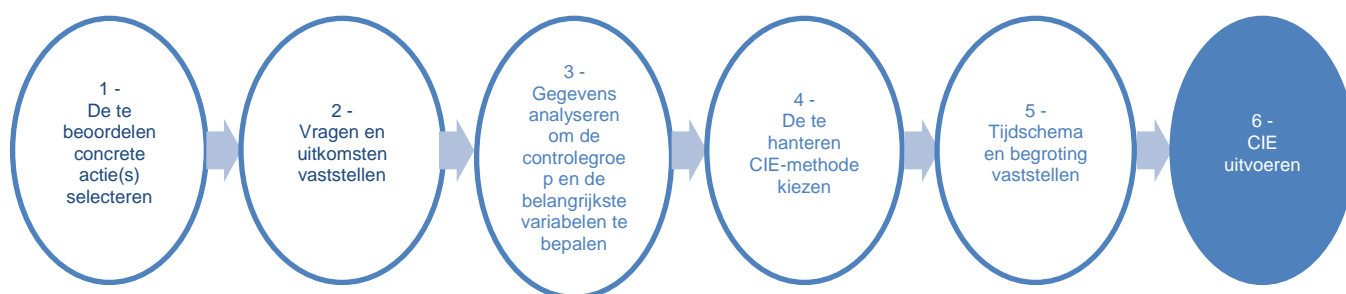
*Focussen op specifieke punten in de tijd ...*

Zoals besproken in punt 1.5 kan het formuleren van een interventielogica (of theorie van verandering) helpen bepalen wanneer de effecten moeten worden geschat.

Een alternatief voor degenen die een CIE plannen bij gebrek aan een interventielogica (maar ook nuttig voor wie wel op een duidelijke interventielogica kan rekenen) is het uitvoeren van een korte evaluatie van eerdere studies waarin interventies worden beoordeeld die vergelijkbaar zijn met de interventie in kwestie. Zorgvuldig onderzoek van de resultaten van eerdere studies kan een goede indicatie geven voor de juiste meting van resultaten en de berekening van effecten.

*...of evaluaties van andere recente studies*

## 2.6. Uitvoering van het CIE



Na de vaststelling van het tijdschema en de begroting is de planning van het CIE afgerond en kan worden overgegaan tot de uitvoering ervan. De laatste stap in deze leidraad is bedoeld om de belangrijkste activiteiten bij de uitvoering van de evaluatie toe te lichten en te laten zien hoe de beheerautoriteit mogelijke belemmeringen kan overwinnen die zij bij de uitvoering ervan kan tegenkomen. In dit deel wordt met name ingegaan op de lessen die zijn getrokken uit de evaluatiehulpdesk, via welke de EC en het Directoraat-generaal Werkgelegenheid bijstand verlenen aan de beheerautoriteiten van alle lidstaten. De uitvoering van de evaluatie wordt in het algemeen besproken in “EVALSED: The Resource for the Evaluation of Socio-Economic Development”<sup>68</sup>; in deze leidraad worden vier hoofdactiviteiten behandeld, met name vanuit het oogpunt van CIE:

1. selectie van de beoordelaar,
2. toezicht op de uitvoering van het CIE,
3. rapportage,
4. benutting van de resultaten.

### 2.6.1. Selectie van de beoordelaar

Ten eerste moet bij de selectie van een beoordelaar worden besloten of een interne dan wel een externe kandidaat zal worden ingeschakeld. In het laatste geval vindt de selectie van de beoordelaar plaats via een openbare aanbestedingsprocedure.

In artikel 44, lid 3, van de GB-verordening wordt erkend dat er sprake kan zijn van een onafhankelijkheidsprobleem en is het volgende bepaald: “De evaluaties worden verricht door interne of externe deskundigen die functioneel onafhankelijk zijn”. Het eerste criterium voor de selectie van een beoordelaar is dus de daadwerkelijke onafhankelijkheid van deze beoordelaar ten opzichte van de beheerautoriteit en de besluitvormingsprocessen van het ESF+-programma. Het tweede criterium is of de nodige vaardigheden en middelen voor de uitvoering van een CIE al dan niet intern beschikbaar zijn. Dit is waarschijnlijk de meest voorkomende belemmering voor interne evaluaties, omdat de technische kennis die nodig is om een CIE uit te voeren zeer specifiek is en de meeste overheidsdiensten hier zelden over beschikken.

Nadat de keuze tussen een interne of externe beoordelaar is gemaakt, moet de overheidsdienst de technische specificaties van de evaluatie vermelden; wanneer er een externe beoordelaar wordt geselecteerd, zullen de

*De onafhankelijkheid van de beoordelaar*

*Belangrijkste elementen van de*

<sup>68</sup> Beschikbaar op: [https://ec.europa.eu/regional\\_policy/sources/docgener/evaluation/guide/guide\\_evalsed.pdf](https://ec.europa.eu/regional_policy/sources/docgener/evaluation/guide/guide_evalsed.pdf)

specificaties deel uitmaken van de openbare aanbestedingsprocedure. In de beschrijving van de opdracht voor een CIE moet speciale aandacht aan de volgende aspecten worden besteed:

*beschrijving van de opdracht*

- Het **doel van de evaluatie** en de onderzoeksvraagstukken moet duidelijk zijn en consequent leiden tot een effectonderzoek en, waar mogelijk, een CIE.
- De **beschikbare gegevens** moeten worden gespecificeerd; dit is een van de belangrijkste problemen die moeten worden aangepakt bij het opzetten en uitvoeren van een CIE. Zo worden beoordelaren zich bewust van de bestaande problemen met betrekking tot de beschikbaarheid van gegevens en kunnen zij realistische en nauwkeurige voorstellen opstellen.
- **Het verzoek om een CIE uit te voeren** kan expliciet in de beschrijving van de opdracht worden opgenomen, maar de keuze van de te gebruiken specifieke methoden moet aan de beoordelaar worden overgelaten. Op die manier kan deze originele oplossingen voorstellen en zijn/haar capaciteit aantonen.
- **Het voorgestelde team** moet over deskundigheid op verschillende gebieden beschikken, met name wat betreft: evaluatie en CIE (weten hoe een evaluatie moet worden opgezet en uitgevoerd), econometrie (met behulp van de juiste methoden), en het beleidsterrein van de interventie in kwestie (reconstrueren van de interventielogica en volledige interpretatie van de bevindingen).
- In **de gunningscriteria** moet voorrang worden gegeven aan de kwaliteit boven de kosten van het voorstel. De kwaliteit van een evaluatie hangt in grote mate af van het vermogen van de beoordelaar om deze op te zetten en uit te voeren. De wettelijke kaders voor overheidsopdrachten zijn vaak niet perfect afgestemd op de behoeften van aanbestedingen voor evaluaties, maar de opdrachtgever zou de bestaande regels zoveel mogelijk moeten aanpassen zodat kwaliteit en technische capaciteit worden beloond.
- Bij **de selectieprocedure** moeten deskundigen op het gebied van evaluatie en CIE in het selectiecomité worden opgenomen om de kwaliteit van de voorstellen ten volle te beoordelen en de verschillende methodologische en organisatorische oplossingen ervan correct te beoordelen.

In sommige lidstaten is de ervaring met CIE's nog steeds minimaal. Dit kan een probleem worden tijdens de selectieprocedure omdat de markt voor deze dienst te beperkt is. Het is mogelijk dat zeer weinig bedrijven of deskundigen kunnen voldoen aan de technische voorwaarden van de opdracht en een voorstel van goede kwaliteit kunnen indienen. In deze gevallen kunnen open seminars van de beheerautoriteit met academici en bedrijven die werkzaam zijn op het gebied van evaluaties helpen bij de voorbereiding van het CIE en de daaropvolgende oproep tot het indienen van voorstellen. De seminars zouden suggesties van nationale deskundigen verzamelen en hen de tijd geven om voor deelname aan de aanbesteding te zorgen; om de transparantie te eerbiedigen en belangenconflicten te vermijden, moeten seminars openbaar zijn en moet basisinformatie ook beschikbaar worden gesteld aan degenen die niet konden deelnemen.

### 2.6.2. Toezicht op CIE's

Na de keuze van de beoordelaar en de start van het CIE moet de beheerautoriteit toezicht houden op het evaluatieproces. Dit heeft voornamelijk betrekking op het controleren en valideren van de prestaties die de beoordelaar moet leveren, het organiseren van voortgangvergaderingen met de beoordelaar en belanghebbenden en, indien nodig, het sluiten van overeenkomsten met gegevenseigenaren.

De beheerautoriteit moet in staat zijn deze taken uit te voeren. Dit betekent ook dat zij de middelen moet hebben om toezicht te houden op het hele proces, en de vaardigheden om technische oplossingen van het CIE te controleren. In het algemeen wijst de beheerautoriteit specifiek personeel toe aan de evaluaties en, wanneer dat personeel onvoldoende ervaring heeft met CIE's, moeten onafhankelijke externe deskundigen ondersteuning bieden. Deze deskundigen kunnen worden geworven op basis van tijdelijke contracten of, wanneer zij afkomstig zijn van openbare instellingen, kosteloos hun diensten leveren. De inspanning zal waarschijnlijk 6 tot 12 werkdagen vergen, wat voldoende zou moeten zijn om rapporten te evalueren en enkele vergaderingen met de beheerautoriteit, beoordelaar en belangrijkste belanghebbenden bij te wonen.

*Interne en externe vaardigheden in het toezichtproces*

Vaak wordt er een stuurgroep opgericht om toezicht te houden op de evaluatieprocessen van een programma. Deze groep bestaat doorgaans uit de beheerautoriteit, ambtenaren die andere uitvoerende afdelingen vertegenwoordigen, de belangrijkste belanghebbenden (sociale partners, ngo's, andere overheidsdiensten) en enkele deskundigen uit de academische wereld en/of overheidsinstellingen (bv. het nationale bureau voor de statistiek). De stuurgroep neemt deel aan de definitie van de evaluatievragen en bespreekt de bevindingen van de evaluatie, de verspreiding en het gebruik ervan in beleidsvormingsprocessen. Een meer flexibele en gespecialiseerde subgroep zou ook samen met de beheerautoriteit toezicht kunnen houden op de evaluatieresultaten.

### 2.6.3. Rapportage

Rapportage is een kernactiviteit van de evaluatie en is het belangrijkste middel om de resultaten bekend te maken. Evaluatieverslagen moeten duidelijk, beknopt en begrijpelijk zijn voor niet-technici en transparant zijn in hun oordelen en beleidsaanbevelingen. Ook moeten zij de betrouwbaarheid van de evaluatie aantonen door de gebruikte methodologie en gegevens te specificeren en de mogelijke grenzen van het onderzoek vast te stellen.

In het algemeen zijn er drie verslagen per CIE:

- Het **aanloopverslag** wordt gewoonlijk kort na ondertekening van het contract ingediend en bevat de definitieve bepaling van methodologie op basis van de met de beheerautoriteit gevoerde besprekingen en de eerste screening van de beschikbare gegevens.
- Het **tussentijdse verslag** wordt ingediend in een tussenliggende fase van de evaluatie en kan in een CIE zijn gewijd aan het tonen van de verzamelde gegevens, de samenstelling van de behandelingsgroep en de controlegroep en de bemonstering.
- Het **eindverslag** bevat de analyses en bevindingen van het CIE. Het bevat nadere informatie over methodologie en gegevens en belicht eventuele beleidsimplicaties die uit de empirische gegevens voortvloeien. De

*Afzonderlijke verslagen om de voortgang te volgen en de resultaten te benutten*

structuur van het eindverslag van een CIE moet enkele belangrijke onderdelen omvatten: evaluatievragen; de interventielogica en vaststelling van de te beoordelen uitkomsten; de gebruikte methodologie; de gebruikte gegevens en de kenmerken van de behandelingsgroep en de controlegroep; de geschatte effecten; antwoorden op de evaluatievragen en hun beleidsimplicaties.

De opdrachtgever beoordeelt alle verslagen waar nodig met de hulp van externe deskundigen, en eventuele opmerkingen worden besproken met de beoordelaar, die de gevraagde verduidelijkingen en verbeteringen zal aanbrengen voordat het verslag formeel wordt goedgekeurd.

In een CIE-verslag moeten overmatige technische aspecten worden vermeden en moeten de bevindingen duidelijk en begrijpelijk zijn voor belanghebbenden en niet-deskundigen. Het hoofdverslag moet een heldere en vloeiende uiteenzetting bevatten. Daarom is het aangewezen om de meer technische stappen van de analyse in speciale bijlagen op te nemen. Details op het gebied van methodologie zijn echter van cruciaal belang om de betrouwbaarheid van het CIE aan te tonen en moeten ook transparant zijn. De identificatie van de uitkomstvariabelen, de gegevenskwaliteit, de samenstelling van de behandelingsgroep en de controlegroep, de steekproefgrootte, de schattingsmethode van de effecten en de gebruikte statistische tests moeten allemaal in het verslag worden opgenomen. Daarnaast zouden, volgens de academische praktijken, de gebruikte geanonimiseerde gegevens en de methoden voor gegevensverwerking toegankelijk kunnen worden gemaakt voor onderzoekers die van plan zijn de analyses te herhalen en te verifiëren.

In het eindverslag van elk CIE moet een gestandaardiseerde reeks minimaal noodzakelijke gegevens worden verstrekt. Deze minimuminformatie heeft een tweeledig doel: 1) een vergelijking van bevindingen afkomstig uit verschillende operationele programma's en verschillende landen vergemakkelijken; 2) informatie verzamelen die beschikbaar zal zijn voor meta-analyse in de toekomst.

*In het eindverslag van een CIE moet minimaal noodzakelijke informatie worden verstrekt*

Het model voor het verzamelen van gestandaardiseerde minimaal noodzakelijke informatie over elke ESF-evaluatie volgt de aanbevelingen in de "Pilot and feasibility study [...]". Vanuit operationeel oogpunt kan deze informatie in een bijlage bij het eindverslag worden opgenomen en ook aan de EC worden toegezonden. Tabel 4 hieronder bevat een mogelijk model voor het verzamelen van de minimaal noodzakelijke informatie.



**Tabel 4 Basisinformatie die moet worden opgenomen in een blad waarin het CIE wordt gepresenteerd**

Informatiecategorie	Specifieke informatie die moet worden verzameld en gerapporteerd
1. Informatie over de interventie	<ul style="list-style-type: none"> <li>- Naam van de geëvalueerde interventie</li> <li>- Kenmerken van de interventie (opleiding, diensten voor arbeidsvoorziening, diensten voor sociale inclusie enz.)</li> <li>- ESF+-programmabebanding (prioritaire assen, specifieke doelstelling, actie)</li> <li>- Beheerautoriteit en/of andere uitvoerende instellingen</li> <li>- Uitgaven</li> <li>- Duur/intensiteit van de door de interventie geleverde behandeling</li> <li>- Deelnemersgroep naar leeftijd, geslacht en geschiktheid (personen met een werkloosheidsuitkering, langdurig werklozen, NEET's)</li> <li>- Territoriale reikwijdte van de interventie</li> </ul>
2. Maatstaf van effectiviteit van de interventie	<ul style="list-style-type: none"> <li>- Indicator die aangeeft of het CIE een positief en statistisch significant effect van de behandeling op de betreffende uitkomst(en) inschat</li> <li>- De werkelijke omvang van dat effect</li> <li>- De twee bovenstaande maatstaven (+/-) en (omvang) voor de reeks relevante en vergelijkbare uitkomsten bv. arbeidsparticipatie, inkomen enz.</li> <li>- Indicatoren van "prijs-kwaliteitverhouding" (indien geschat)</li> </ul>
3. Gegevens en methodologie	<ul style="list-style-type: none"> <li>- Gegevensbron (enquête, administratieve gegevens)</li> <li>- Observatieperiode</li> <li>- Econometrische CIE-methode</li> <li>- Tijdshorizon (doorgaans als volgt: kort, nl. maximaal 12 maanden vanaf het einde van de interventie, middellang, nl. 12 tot maximaal 24 maanden, langetermijneffecten, nl. meer dan 24 maanden)</li> <li>- Steekproefgrootte van de behandelingsgroep en de controlegroep</li> <li>- Beschikbaarheid/opname van gegevens van vóór het programma</li> </ul>

#### 2.6.4. Benutting van de resultaten

De benutting van het resultaat is belangrijk omdat de evaluatie weinig effect zal hebben wanneer de resultaten niet op doeltreffende wijze worden verspreid en het beoogde publiek niet bereiken.

*Verspreiding van de bevindingen van het CIE is van cruciaal belang*

De verspreiding van de bevindingen en outputs van de evaluatie omvat normaal gesproken:

- ten minste één schriftelijk evaluatieverslag, met inbegrip van een kort overzicht en een samenvatting;
- ten minste één mondelinge presentatie van de bevindingen, ondersteund door dia's of soortgelijke hulpmiddelen;
- een technisch gedeelte van het verslag (of een technische bijlage) met een nauwkeurige beschrijving van de gebruikte methodologie, de belangrijkste aannames en de methode die bij de statistische analyses is gevolgd.

Alle evaluatieverslagen moeten openbaar worden gemaakt. Dit is een bepaling in de verordening houdende gemeenschappelijke bepalingen voor de programmeringsperiode 2021-2027<sup>69</sup>. Daarnaast moeten de evaluatiebevindingen worden gepresenteerd en besproken in het monitoringcomité van het programma; dit is een verplichte stap, die op zich echter niet voldoet om een diepgaand debat aan te zwengelen, omdat operationele prioriteiten doorgaans voorrang hebben tijdens de

<sup>69</sup> Artikel 44, lid 7, GB-verordening.

vergaderingen van de monitoringcomités. Zie onderstaand tekstkader voor een voorbeeld van een nationale praktijk voor de presentatie en verspreiding van de evaluatiebevindingen.

Er bestaat niet één enkel doeltreffend communicatiebeleid om de bevindingen te verspreiden en een debat over de getrokken lessen op gang te brengen; elke context vereist een eigen strategie. Het is dan ook van belang met een doeltreffende strategie te komen en er met name voor te zorgen dat belanghebbenden buiten de beheerautoriteit in kennis worden gesteld van de bevindingen. Het is mogelijk om het “publiek” van het CIE in hoofdgroepen te verdelen en de belangrijkste soorten communicatie vast te stellen. Bijvoorbeeld:

- beleidsmakers op hoog niveau en politieke vertegenwoordigers: zij zijn het gewoon korte documenten (beleidsnota’s, samenvattingen) met de belangrijkste resultaten en aanbevelingen te lezen;
- ambtenaren van de instantie die betrokken is bij de uitvoering van de interventie of soortgelijke beleidsmaatregelen: seminars of workshops kunnen een doeltreffend instrument zijn om CIE-resultaten te presenteren, waarbij het technische niveau van de presentatie moet worden afgestemd op het type interactie met de aanwezigen;
- belanghebbenden bij het beleid (sociale partners, ngo’s, begunstigen enz.): samenvattingen van de verslagen in combinatie met jaarlijkse vergaderingen over de bevindingen van de evaluatie kunnen het juiste instrument zijn om deze actoren bij het debat te betrekken. Met name wanneer deze actoren hebben bijgedragen aan de vaststelling van de evaluatievragen, moet de rapportage van de resultaten van de evaluatie aan die vragen worden gekoppeld;
- deskundigen en academici: wetenschappelijke conferenties of seminars kunnen de beste plaatsen zijn om de resultaten van de evaluatie te presenteren. Deze actoren valideren de resultaten van de evaluatie vanuit wetenschappelijk oogpunt en versterken zo de betrouwbaarheid ervan.

*Veel soorten publiek en vele communicatiemethoden*

Een meta-evaluatie<sup>70</sup> is een krachtig aanvullend instrument voor de verspreiding van de resultaten van CIE’s. Dit instrument maakt bevestiging en generalisatie mogelijk van de belangrijkste bevindingen die één CIE alleen niet kan opleveren. Om een meta-evaluatie uit te voeren, moet men beschikken over een groot aantal betrouwbare en kwalitatief goede CIE’s inzake vergelijkbaar beleid; op dit moment zijn meta-evaluaties nog steeds zeldzaam en niet uitsluitend gericht op ESF-interventies. Meta-evaluaties zouden beter ingeburgerd geraken als er breder gebruik zou worden gemaakt van CIE’s en hun planning vooraf zodat een vergelijking achteraf gemakkelijker wordt. In de programmeringsperiode 2021-2027 van het ESF+ kunnen meta-evaluaties worden uitgevoerd in landen met een aanzienlijk aantal programma’s of door middel van vergelijkingen tussen landen.

*Meta-evaluaties voor het versterken van verspreiding en generalisatie*

<sup>70</sup> Dit is een systematische vergelijking, ook met behulp van statistische instrumenten, van talrijke CIE’s van vergelijkbare interventies. Een meta-evaluatie geeft aanwijzingen over de doeltreffendheid van een type interventie door de effecten van soortgelijke maatregelen in veel verschillende contexten te onderzoeken. Meta-evaluaties kunnen echter niet ingaan op de details van elke afzonderlijke interventie die in de vergelijking wordt gebruikt, en hun beleidsimplicaties hebben betrekking op brede beleidslijnen en niet op afzonderlijke details voor elke interventie.

### Tekstkader 13 De Poolse ervaring met evaluatieconferenties

De *International Evaluation Conference*, georganiseerd door het ministerie van Ontwikkelingsfondsen en Regionaal Beleid en het Poolse agentschap voor bedrijfsontwikkeling, is een van de belangrijkste elementen op het gebied van evaluatie van overheidsinterventies in Polen. Sinds haar oprichting is de conferentie een vast platform geworden voor discussie tussen de belanghebbenden die verantwoordelijk zijn voor beleidsvorming, beleidsuitvoering en effectbeoordeling.

De conferentie ging van start in 2005 en werd tien jaar lang jaarlijks en sinds 2015 elke twee jaar gehouden, waardoor het evenement van 2021 de 14<sup>e</sup> editie was. Momenteel is de conferentie over twee dagen verspreid. Eén dag is gewijd aan onderzoeksresultaten en effecten van de interventies; dit is de dag waarop beleidskwesties worden besproken. De tweede dag is gericht op methodologische kwesties en de evaluatie als proces.

Tijdens de 14 edities van de conferentie zijn contrafeitelijke methoden meerdere keren het onderwerp geweest van paneldiscussies of presentaties van deskundigen, en werden zij op beide dagen vanuit verschillende invalshoeken benaderd. Zo werden in 2017 de geschatte effecten gepresenteerd tijdens sessies die betrekking hadden op de impact van de interventies, terwijl er afzonderlijke besprekingen plaatsvonden over de voor- en nadelen van schattingstechnieken.

De conferentie is ook van belang vanwege de diversiteit van het publiek, aangezien zij niet alleen gericht is op onderzoekers of specialisten op het gebied van methodologie, maar ook op degenen die interventies opzetten en uitvoeren. De doelstellingen van het evenement zijn tweeledig: de capaciteit van het evaluatiesysteem opbouwen door de vaardigheden en kennis van de leden ervan te ontwikkelen, en netwerken vormen door belanghebbenden de gelegenheid te geven om ervaringen uit te wisselen met nationale, regionale en buitenlandse partners. Elke editie brengt ongeveer 300 tot 400 vertegenwoordigers uit verschillende sectoren bijeen: de overheid, de academische wereld, adviesbureaus en ngo's. Dit is een aanzienlijk aantal, ook vanwege het feit dat de empirisch onderbouwde beleidsaanpak in Polen nog niet volledig wortel heeft geschoten.

Deze internationale evaluatieconferentie maakt deel uit van het evaluatiesysteem met betrekking tot overheidsdiensten. Aanzienlijke capaciteitsopbouw was mogelijk dankzij opleidingen die werden verzorgd door het Gemeenschappelijk Centrum voor onderzoek en de Wereldbank, of die werden gefinancierd in het kader van technische bijstand in operationele programma's. Opleiding op het gebied van evaluatie, met inbegrip van de contrafeitelijke methode, was niet alleen beschikbaar voor werknemers van evaluatie-eenheden bij beheerautoriteiten, maar ook voor instellingen zoals lokale of regionale arbeidsbureaus. De conferentie is een gelegenheid om actuele ontwikkelingen en bevindingen te verzamelen, te bespreken en te delen<sup>71</sup>.

<sup>71</sup> Informatie over de conferentie is te vinden op: <https://www.ewaluacja.gov.pl/strony/xiv-miedzynarodowa-konferencja-ewaluacyjna/>. Met dank aan Piotr Strzęboszewski en andere ambtenaren van het ministerie van Ontwikkelingsfondsen en Regionaal Beleid voor de informatie die zij over het conferentiesysteem hebben verstrekt.

# Hoofdstuk 3. Hoe kan de juiste methode voor de uitvoering van een CIE worden gekozen?

In dit hoofdstuk worden empirische methoden voor contrafeitelijk effectonderzoek voorgesteld. Met name wordt de experimentele methode (gerandomiseerd onderzoek met een controlegroep) besproken, evenals de meest voorkomende quasi-experimentele methoden: matching (op basis van de “propensity-score”), difference-in-differences, regressie-discontinuïteitsanalyse en instrumentele variabelen. Elke methode wordt gekenmerkt door een specifieke manier om een controlegroep te genereren om de contrafeitelijke vraag te beantwoorden: “Wat zou er met de behandelingsgroep zijn gebeurd qua uitkomsten als deze niet aan de interventie had deelgenomen/was onderworpen?”. In de praktijk kan de gekozen opzet worden afgestemd op de betreffende context, afhankelijk van het type interventie en de beschikbare of te verzamelen gegevens<sup>72</sup>.

Het is niet mogelijk gedetailleerde richtsnoeren te geven voor de keuze van de meest geschikte evaluatieopzet in de zeer uiteenlopende omstandigheden waarmee beheerautoriteiten te maken hebben. Bij het kiezen van de meest relevante benadering van CIE's in een bepaalde context, moeten de beheerautoriteiten rekening houden met wat bij eerdere evaluaties goed heeft gewerkt, zowel binnen de beheerautoriteit zelf, binnen de lidstaat als in andere lidstaten. Beheerautoriteiten kunnen leren van wat eerder is bereikt binnen hun programma en erbuiten, waar sprake was van vergelijkbare omstandigheden. In dit verband kunnen fora voor de uitwisseling van geleerde lessen bij de opzet en uitvoering van evaluaties nuttige informatiebronnen zijn. Ook in de literatuur zoeken naar evaluaties van soortgelijke interventies kan belangrijke informatiebronnen opleveren die kunnen helpen bij de opzet. Deskundigen die door de beheerautoriteit zijn aangesteld, zullen eveneens met ideeën komen over hoe de opzet van een evaluatie het best kan worden aangepakt. Daarnaast mag niet worden vergeten dat er binnen de beheerautoriteiten aanzienlijke expertise en ervaring kunnen bestaan die verder kunnen worden benut.

*Selectie van de juiste methode*

## 3.1. Randomisatie — de experimentele benadering

### *Belangrijkste kenmerken van randomisatie*

- Personen die voor deelname in aanmerking komen, worden willekeurig toegewezen aan een behandelings- of een controlegroep.
- Randomisatie zorgt ervoor dat beide groepen (gemiddeld) identiek zijn wat alle relevante kenmerken betreft.

<sup>72</sup> Zie voor andere technische methodologische aspecten ook [Europese Commissie, 2019](#)

- De controlegroep geeft derhalve het antwoord op de contrafeitelijke vraag, en het verschil in de uitkomsten tussen de behandelingsgroep en de controlegroep geeft het causale effect van de interventie aan.

Gerandomiseerde evaluaties kunnen verschillende vormen aannemen. Hier wordt een eenvoudige aanpak met twee groepen gehanteerd — een behandelingsgroep en een controlegroep — om de belangrijkste beginselen te verduidelijken. In figuur 1 is een eenvoudige gerandomiseerde opzet weergegeven.

Het belangrijkste punt is dat randomisatie ervoor zorgt dat de twee groepen op het moment van randomisatie in alle opzichten statistisch gelijkwaardig zijn. Vervolgens ondergaat de behandelingsgroep de interventie waarop de evaluatie is gericht en waarvan de gevolgen of effecten moeten worden gemeten.

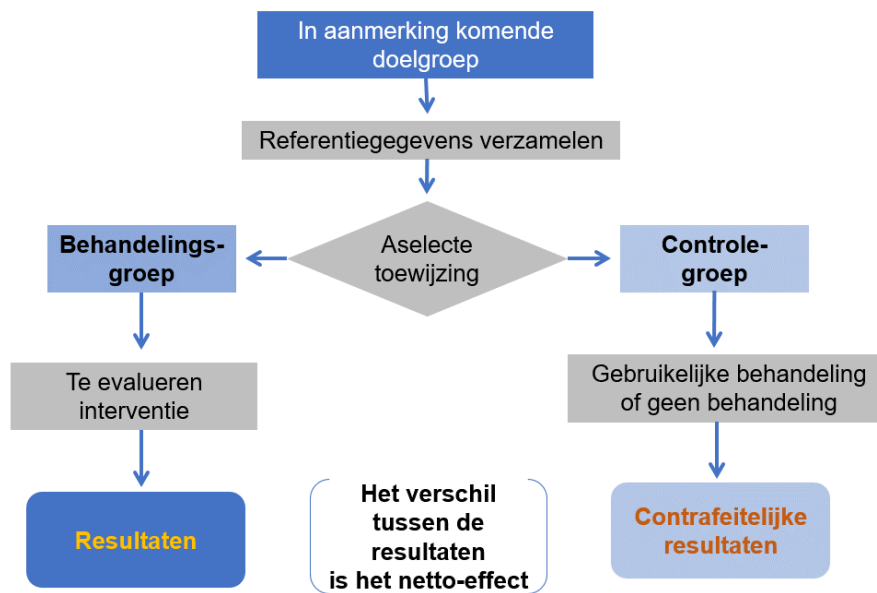
*Statistisch  
gelijkwaardige  
groepen*

Afhankelijk van de centrale beleidsvraag, kan er een controlegroep worden toegewezen die geen enkele behandeling ondergaat, of kan de behandelingsgroep worden vergeleken met een groep die is onderworpen aan een andere relevante behandeling (die kan worden opgevat als representatief voor een gebruikelijke behandeling); ook kunnen er ten aanzien van een controlegroep meerdere behandelingsgroepen zijn. Het kan bijvoorbeeld interessant zijn om de effecten van een uit het ESF gefinancierd opleidingsprogramma te vergelijken met andere op nationaal niveau gefinancierde opleidingen, of met de verstrekking van andere diensten aan dezelfde groepen.

*Geen behandeling  
of een andere  
behandeling voor  
de controlegroep*

Aangezien de behandelingsgroep en de controlegroep statistisch gelijkwaardig zijn op het moment van randomisatie en de onderwerping aan daaropvolgende behandelingen wordt gecontroleerd, kunnen verschillen in resultaten worden toegeschreven aan de interventie die wordt geëvalueerd (met de gebruikelijke statistische onzekerheden) en kunnen alternatieve verklaringen worden uitgesloten als de oorzaken van eventuele waargenomen verschillen (zie het voorbeeld in het tekstkader hieronder).

**Figuur 6 Gerandomiseerd gecontroleerd onderzoek met twee groepen**



Door de intrinsieke kenmerken van hun opzet en op voorwaarde dat zij correct worden uitgevoerd, maken gerandomiseerde modellen het mogelijk krachtige bewijzen te leveren voor de effecten van een interventie. Om deze reden genieten zij sterk de voorkeur. Zij vereisen echter een vroegtijdige en gedetailleerde planning en kunnen lastig zijn om op te zetten en te beheren. Bovendien hebben programmabeheerders te maken met grote uitdagingen bij de correcte uitvoering ervan. Zo kan het randomisatieproces zelf de samenstelling van de deelnemers aan een interventie veranderen: sommige potentiële deelnemers kunnen namelijk worden afgeschrikt door het randomisatieconcept en weigeren deel te nemen. Daarnaast voldoen personen die aan randomisatie worden onderworpen mogelijk niet altijd aan de toewijzingsstatus en zijn er een reeks andere uitdagingen die eventueel moeten worden aangepakt. In bepaalde omstandigheden kan de uitvoering van een gerandomiseerd onderzoek met een controlegroep duur zijn.

*Sterke bewijzen...*

Om deze en andere redenen lijkt het onwaarschijnlijk dat evaluaties van uit het ESF gefinancierde instrumenten en interventies op een gerandomiseerde methode zullen berusten. In deze leidraad wordt echter gewaarschuwd voor de neiging om randomisatie zonder zorgvuldige overweging in alle gevallen uit te sluiten. De benadering is op grote schaal toegepast en de talrijke voorbeelden omvatten onder meer de GAIN-experimenten uit de Verenigde Staten<sup>73</sup> (er zijn talrijke andere voorbeelden uit Noord-Amerika), experimenten in Zweden<sup>74</sup> tijdens ESF-interventies in de programmeringsperiode 2014-2020 (zie het volgende tekstkader) en een studie die in Duitsland is uitgevoerd om de effecten te beoordelen van actieve arbeidsmarktdiensten die door particuliere dienstverleners worden verstrekt, in vergelijking met die welke via de openbare dienst voor arbeidsvoorziening worden verleend<sup>75</sup>.

*... maar moeilijk op te zetten*

<sup>73</sup> Zie [Riccio J., Friedlander, D., Freedman S., 1994](#)

<sup>74</sup> Zie [Hagglund, P., 2006](#)

<sup>75</sup> Zie [Krug G., Stephan G., 2011](#)

## Tekstkader 14 Voorbeeld van een gerandomiseerd onderzoek van een ESF-project voor jongeren

### De evaluatie van het project “Ung framtid” (“Jonge toekomst”) in Zweden

Het Zweedse initiatief “Ung framtid” (“Jonge toekomst”) is een uit het ESF gefinancierd project dat tussen 2015 en 2018 werd uitgevoerd. Het was bedoeld om jonge werkzoekenden tussen 18 en 24 jaar die de arbeidsmarkt betreden in Midden-Norrland, noordelijk Midden-Zweden en Zuid-Zweden te ondersteunen. Het project intensiverde en personaliseerde ondersteuningsactiviteiten gericht op jonge werkzoekenden en werd uitgevoerd door *Arbetsformedlingen*, de Zweedse openbare dienst voor arbeidsvoorziening. Het bestond uit individuele planning, verstrekking van informatie, advies en concrete ondersteuning op het gebied van matching, opleiding en andere mogelijke activiteiten. In totaal waren bijna 17 000 jongeren bij het project betrokken.

De evaluatie werd uitgevoerd in opdracht van de Zweedse ESF-Raad bij *Arbetsformedlingen*, als onderdeel van het evaluatiekader dat was ontwikkeld binnen een ESF-project voor de periode 2014-2020 ter verbetering van de evaluatiecapaciteit van de ODA's (empirisch onderbouwd EU-financieringsproject 2014-2020; zie het tekstkader in het laatste hoofdstuk van de leidraad).

In een experiment dat tussen juni 2017 en januari 2018 werd uitgevoerd met acht plaatselijke arbeidsbureaus op een totaal van 90 die deelnamen aan het project, werden in aanmerking komende jongeren willekeurig toegewezen aan de behandelingsgroep (die intensieve ondersteuning kreeg) en de controlegroep (die normale ondersteuning kreeg)<sup>76</sup>. In het experiment werden 4 689 jongeren willekeurig toegewezen aan de behandelingsgroep en de controlegroep, waarbij er 2 972 aan de controlegroep werden toegewezen. Het aselechte toewijzingsproces leidde tot een behandelingsgroep en een controlegroep die op het moment van toewijzing zeer vergelijkbaar waren. Daarom zijn eventuele verschillen tussen de twee groepen met betrekking tot de uitkomstvariabelen (uitstroom uit de werkloosheid, aandeel van werklozen en gemiddeld aantal dagen werkloosheid), gemeten na toetreding tot de ESF-interventie, toe te schrijven aan de “Ung framtid”-activiteiten.

Uit de bevindingen van de studie blijkt dat het project “Ung framtid” positieve effecten had voor vrouwen, maar niet voor mannen. Voor vrouwen werden positieve effecten aangetroffen in alle onderzochte uitkomstvariabelen: de uitstroom uit de werkloosheid van behandelde vrouwen lag in de eerste twee maanden na de toetreding tot het project ongeveer zeven procentpunten hoger, het percentage werklozen nam sterker af in de behandelingsgroep dan in de controlegroep, en de behandelde vrouwen waren minder dagen werkloos dan de vrouwen in de controlegroep<sup>77</sup>.

Gerandomiseerde opzetten onderscheiden zich voornamelijk van andere methoden omdat zij sterk gericht zijn op het beheersen van potentiële vertekening tussen de behandelings- en de controlegroep door middel van de onderzoeksopzet. Door deze nadruk op de opzet is deze methode zeer intuïtief, maar zij vereist wel planning vooraf. Gerandomiseerde opzetten worden vaak het best toegepast bij de evaluatie van nieuwe proefprojecten in plaats van lopende projecten. De reden hiervoor is dat er enige mate van controle is vereist met betrekking tot de wijze waarop deelnemers worden aangeworven voor de te evalueren interventie. Deze “controle” is vaak moeilijker te verkrijgen in bestaande programma's dan bij nieuwe interventies die openstaan voor nieuwe ideeën.

*Randomisatie door onderzoeksopzet*

Het uitvoeren van een gerandomiseerd onderzoek met een controlegroep houdt in dat een deel van de in aanmerking komende doelgroep niet wordt uitgenodigd om deel te nemen, maar aan een controlegroep wordt toegewezen. Deze toewijzing is volledig willekeurig en vindt niet plaats op verzoek van de aanvrager of de beheerders van de interventie. Daarom

*Ethische bezwaren*

<sup>76</sup> In april en mei 2017 werd een proeffase uitgevoerd om de verschillende activiteiten te testen, zoals het beheer van de randomisatie, de betrokkenheid van de jonge deelnemers, de rapportage.

<sup>77</sup> Axdorph E, Egebark J., Lundström T., Özcan G., 2019.

kunnen beleidsmakers geneigd zijn om op ethische gronden bezwaar te maken tegen gerandomiseerde onderzoeken met een controlegroep, nog voordat zij overwegen of deze vanuit praktisch en analytisch oogpunt haalbaar zijn.

Toch is er veel te zeggen voor gerandomiseerde opzetten. Als randomisatie kwalitatief gezien het beste en het betrouwbaarste bewijs van de doeltreffendheid van door de overheid gefinancierde interventies oplevert, is het belangrijk dat deze opzet op grotere schaal wordt toegepast bij de beoordeling van de effecten van de interventie. Bovendien is het, als de effecten van een bepaalde interventie a priori onbekend zijn, niet onethisch om personen uit te sluiten, aangezien niet kan worden aangenomen dat zij hier enig voordeel van zouden hebben verkregen. Daarnaast worden dergelijke benaderingen op grote schaal gebruikt in de geneeskunde en in andere vakgebieden, zoals onderwijsonderzoek. Tot slot is het willekeurig toewijzen aan de interventie van personen uit de pool van diegenen die daarvoor in aanmerking komen in gevallen waarin er voor de door een interventie verleende diensten en ondersteuning te hoog wordt ingetekend (d.w.z. dat er meer personen zijn die in aanmerking komen voor steun dan die welke daadwerkelijk van de interventie kunnen gebruikmaken), de meest ethische wijze om beperkte middelen toe te wijzen.

*... maar er zijn ook sterke argumenten voor het gebruik van een gerandomiseerde opzet*

## 3.2. Niet-gerandomiseerde of quasi-experimentele opzetten

Wanneer randomisatie niet haalbaar is, bestaat er een reeks alternatieve methoden in de CIE-toolbox waarin verschillende benaderingen worden gebruikt om een controlegroep samen te stellen en zo de contrafeitelijke vraag te beantwoorden. In dit onderdeel wordt eerst de belangrijkste uitdaging voor quasi-experimentele methoden in het algemeen uiteengezet en worden vervolgens vier benaderingen gepresenteerd die gebruikelijk zijn in de empirische CIE-praktijk.

*Wanneer randomisatie niet mogelijk is*

### 3.2.1. Doel- en controlegroep zonder randomisatie

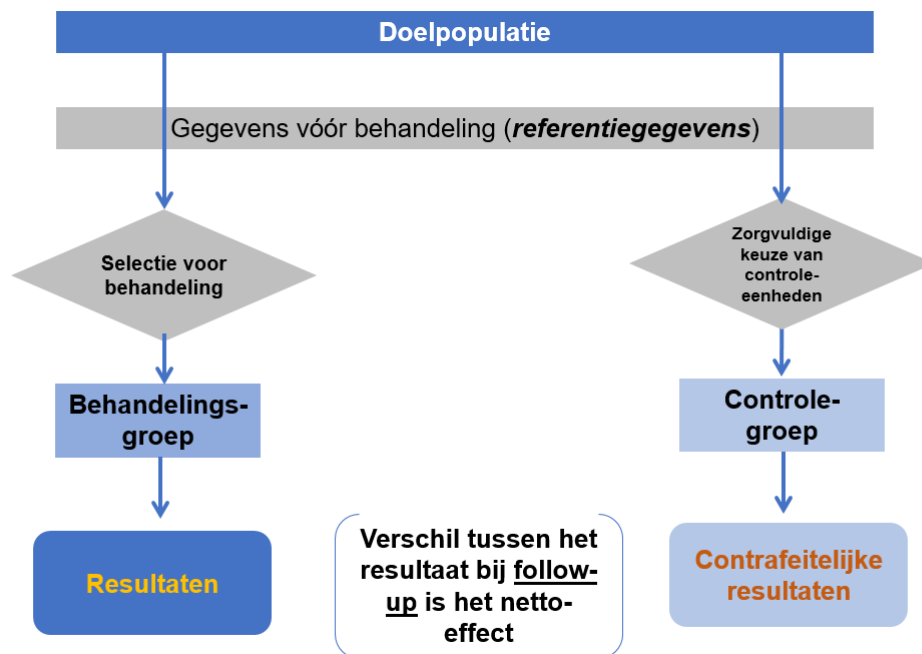
Bij quasi-experimentele opzetten worden de doelgroepen die de interventie ondergaan vergeleken met een controlegroep<sup>78</sup> van niet willekeurig toegewezen doelgroepen of potentiële doelgroepen die de interventie niet ondergaan. Net als bij de experimentele methode, is het doel een objectieve schatting te verkrijgen van de verandering in de uitkomsten die de onderzochte interventie heeft teweeggebracht. Aangezien de behandelings- en de controlegroep niet willekeurig worden samengesteld, vereisen quasi-experimentele opzetten veel meer aandacht voor methoden die rekening houden met potentiële verschillen tussen de leden van de behandelingsgroep en de potentiële leden van de controlegroep, die van invloed kunnen zijn op de beslissing om deel te nemen en dus op de resultaten. Het is van fundamenteel belang een plausibele controlegroep te selecteren. Wanneer er geen geschikte controlegroep wordt geselecteerd en de blijvende verschillen tussen de twee groepen in de analyse niet worden verklaard, wordt de geloofwaardigheid van de schattingen verzwakt

<sup>78</sup> Strikt genomen verwijst de term “controlegroep” naar een experimentele context, terwijl in een quasi-experimentele context meestal de term “vergelijkingsgroep” wordt gebruikt. In de praktijk worden de termen echter vaak door elkaar gebruikt. Daarom verwijst in deze leidraad de term “controlegroep” naar de groep die voor het schatten van de contrafeitelijke situatie wordt gebruikt, ongeacht de CIE-opzet.



en kunnen pogingen om alternatieve verklaringen voor waargenomen effecten uit te sluiten, op niets uitlopen.

**Figuur 7 Gestileerde quasi-experimentele opzet met behandelings- en controlegroep**



Wat de uit het ESF medegefinancierde interventies betreft, is de meest gebruikte quasi-experimentele evaluatieopzet een op twee groepen gebaseerde referentie-/follow-up-opzet. Dergelijke opzetten hebben net als bij randomisatie een controlegroep en een behandelingsgroep, met het verschil dat de controlegroep (zonder het gebruik van randomisatie) is samengesteld uit leden van bestaande niet-deelnemende groepen, zodat zij zoveel mogelijk op de behandelde groep lijkt.

*De controle- en de behandelingsgroep moeten vergelijkbaar zijn*

Een goede strategie voor het vinden van een geldige controlegroep binnen een quasi-experimentele context is het selecteren van controlepersonen die van de behandeling zijn uitgesloten op basis van factoren die geen verband houden met hun kenmerken en potentiële resultaten. In sommige omstandigheden kan er reden zijn om aan te nemen dat, hoewel de controlegroepen niet expliciet willekeurig zijn samengesteld, er achteraf personen of ondernemingen kunnen worden geïdentificeerd waarvan het feit dat zij de behandeling niet hebben ondergaan willekeurig blijkt te zijn ten opzichte van de potentiële resultaten. Als deze omstandigheden zich voordoen, is dit een bijna ideale context voor een quasi-experimentele benadering. Zo kunnen bepaalde leden van de doelgroep van een interventie worden uitgesloten van deelname aan de interventie als gevolg van een administratieve nalatigheid of vergissing. Inzicht in het selectieproces voor de behandeling is daarom van groot belang voor het samenstellen van een geldige controlegroep — dit kan niet genoeg worden benadrukt.

Een geloofwaardige controlegroep kan op meerdere manieren worden samengesteld. Ten eerste kan een statistische matchingbenadering worden gevolgd: in dit geval worden gegevens verzameld van zowel behandelde personen als van een (doorgaans zeer grote) steekproef van niet-behandelde personen. Vervolgens wordt een controlegroep samengesteld uit de groep niet-behandelde personen door degenen te kiezen die het vergelijkbaarst zijn

*Het koppelen van behandelde en niet-behandelde personen*

met de personen in de behandelingsgroep. “Vergelijkbaarheid” verwijst naar een reeks sociaal-demografische kenmerken, bijvoorbeeld leeftijd, geslacht, onderwijs, arbeidsparticipatie en werkloosheid enz., gemeten op een tijdstip voordat de behandelingsgroep tot het programma toetrad. Op basis daarvan worden niet-behandelde personen daadwerkelijk “gematcht” met behandelde personen. In de praktijk kan de mogelijk lange lijst van sociaal-demografische kenmerken worden samengevat aan de hand van de “propensity score”-waarde, wat het gebruik ervan vergemakkelijkt.

### 3.2.2. Propensity score matching

#### *Belangrijkste kenmerken van propensity score matching*

- Net als bij randomisatie wordt bij matching achteraf een controlegroep samengesteld die wat alle relevante kenmerken betreft zo vergelijkbaar mogelijk is met de behandelingsgroep.
- In tegenstelling tot randomisatie kunnen alleen waarneembare kenmerken (leeftijd, geslacht, onderwijs enz.) worden vergeleken, terwijl niet-waarneembare kenmerken (zoals motivatie) niet in aanmerking kunnen worden genomen.
- De geldigheid van de methode hangt sterk af van de beschikbaarheid van gegevens.

**Propensity score matching** (PSM) houdt in dat voor de gehele steekproef (behandelingseenheden en potentiële controle-eenheden) een statistisch model wordt geraamd dat een geschatte waarschijnlijkheid van deelname voor elke persoon of onderneming oplevert, ongeacht of deze persoon daadwerkelijk deelneemt of niet<sup>79</sup>. Behandelde personen of ondernemingen worden vervolgens op basis van de propensity-score gematcht met één niet-behandelde persoon of onderneming, of met meerdere niet-behandelde personen of ondernemingen<sup>80</sup>. Aan de hand van deze procedure wordt een controlegroep samengesteld die vervolgens kan worden gebruikt om een schatting van het contrafeitelijke effect te maken. Op deze manier zorgt matching ervoor dat bij de schattingen van de effecten rekening wordt gehouden met de waarneembare verschillen tussen de behandelde groep en de controlegroep, en kan, ervan uitgaande dat alle relevante verschillen voorafgaand aan de matching waarneembaar zijn, een objectieve schatting van de interventie-effecten worden verkregen. Indien de selectie voor behandeling echter op niet-waargenomen factoren gebaseerd is, blijft de vraag bestaan of matching geschikt is om vertekening te kunnen wegnemen. Matching is gebaseerd op de cruciale veronderstelling dat het selectieproces wordt gekenmerkt door waarneembare gegevens.

Onderstaande figuur toont een intuïtieve en vereenvoudigde weergave van de propensity score matching-methode. De y-as staat voor het aantal personen in de behandelde en de niet-behandelde groep, gerangschikt naar de propensity-score op de x-as. Normaal gesproken hebben behandelde

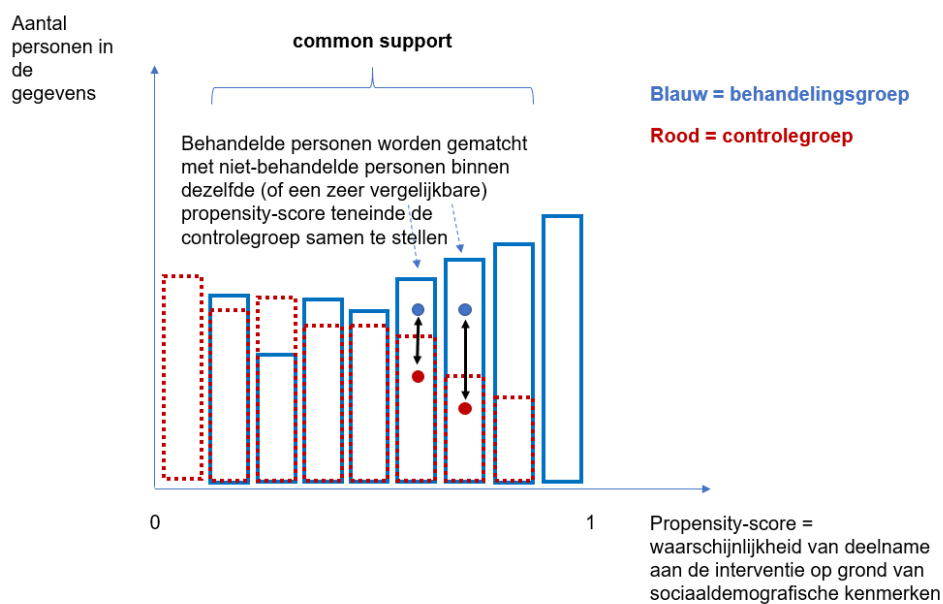
*Waarschijnlijkheid van deelname als manier om de behandelde groep en de controlegroep te definiëren*

<sup>79</sup> Om deze uiteenzetting te vereenvoudigen, wordt ervan uitgegaan dat de beleidsmakers het effect van de behandeling wensen te kennen op degenen die daadwerkelijk diensten van het programma hebben ontvangen (dit is in vele gevallen een subgroep van de doelgroep die deze kans heeft gekregen). Dit wordt een TOT-analyse genoemd, waarbij het effect van de “behandeling op de behandelde” (*treatment on the treated*, TOT) wordt vastgesteld.

<sup>80</sup> Er bestaan heel wat potentiële benaderingen van matching op propensity-score. Zie voor een toegankelijk overzicht [Caliendo M., Kopeinig S., 2008](#)

personen relatief hogere propensity-scores, terwijl die van niet-behandelde personen doorgaans lager liggen. Het gebied waarop de propensity-scores voor beide groepen elkaar overlappen, staat bekend als het gebied van *common support*.<sup>81</sup> In dit gebied worden behandelde gevallen gematcht met niet-behandelde gevallen. In het diagram worden twee voorbeelden gegeven, maar het proces wordt in wezen herhaald totdat elk behandeld geval is gematcht met een niet-behandeld geval binnen het gebied van common support. In de figuur wordt dit gedaan met behulp van “nearest-neighbour” matching. De “nearest neighbour” (naaste buur) van elk lid van de behandelingsgroep is het lid van de controlegroep met de dichtstbijzijnde propensity-score. Zodra twee groepen zijn gevormd, kunnen hun gemiddelde resultaten worden vergeleken om een schatting van het effect te kunnen maken. In de praktijk kan propensity score matching een zeer complex proces worden met een reeks kwesties die in overweging moeten worden genomen. Veel van deze kwesties worden hier niet in aanmerking genomen om ervoor te zorgen dat de kernbeginselen duidelijk zijn. In het volgende tekstkader wordt een ESF-evaluatie beschreven, waarbij gebruik werd gemaakt van een matchingmethode.

**Figuur 8 Weergave van de propensity score-methode**



De plausibiliteit van de propensity score-methode berust onder meer op de aanname dat de selectie voor behandeling volledig kan worden gebaseerd op waarneembare gegevens. Met andere woorden, er zijn geen niet-waargenomen verschillen tussen de behandelings- en de controlegroep die verband houden met de resultaten en/of het besluit om aan de interventie deel te nemen. De plausibiliteit van deze aanname wordt versterkt door de opname van een breed scala van variabelen in de schatting van de propensity-scores, waarbij de selectie van variabelen gebaseerd is op bestaande kennis en theorie. Met name wordt in het kader van arbeidsmarktinterventies de opname van individuele arbeidsmarktsituaties uit het verleden ten eerste

*Selectie op basis van waargenomen gegevens*

<sup>81</sup> De omvang van het gebied van common support heeft gevolgen voor de steekproefgrootte en de bruikbaarheid van de resultaten voor het beleid, vooral wanneer een groot aantal behandelde gevallen buiten het gebied van common support valt.

aanbevolen bij het nagaan of er potentiële niet-waargenomen verschillen bestaan<sup>82</sup>.

### **Tekstkader 15 Voorbeeld van een evaluatie waarbij gebruik wordt gemaakt van een matchingmethode<sup>83</sup>**

#### **De effecten van ESF-interventies die in 2014-2020 zijn gefinancierd voor langdurig werklozen in de regio Marche<sup>84</sup>**

Er werd gebruikgemaakt van een matchingmethode om het effect te beoordelen van een aantal ESF-interventies (stages, werkbeurzen, werkervaring in gemeenten en opleidingsvouchers) op langdurig werklozen, die in de Italiaanse regio Marche werden gefinancierd. De interventies waren niet specifiek gericht op langdurig werklozen, maar zij vormden de meerderheid van de deelnemers aan alle geanalyseerde ESF-interventies.

Het doel van de evaluatie was na te gaan wat het effect is van ESF-maatregelen ten aanzien van de waarschijnlijkheid om na 6, 9, 12, 15 en 18 maanden vanaf het begin van de interventies een baan te vinden. Er werden verschillende uitkomstvariabelen gebruikt: de waarschijnlijkheid om op een bepaald moment na de interventies in dienst te worden genomen, de waarschijnlijkheid in dienst te worden genomen met een overeenkomst voor onbepaalde tijd en het aantal gewerkte dagen in een periode na de interventies.

Om de verschillen tussen de behandelde en de niet-behandelde groep vast te stellen, werd gekozen voor een propensity score-methode (nearest-neighbour matching) en werden de vier hierboven genoemde ESF-interventies afzonderlijk geanalyseerd. Voor elke interventie bestond de behandelde groep uit deelnemers die vóór eind augustus 2019 met de interventie begonnen waren (526 personen met een stage, 1 058 met een werkbeurs, 236 met werkervaring in gemeenten en 241 met opleidingsvouchers); de controlegroep bestond uit werklozen die in de periode 2016-2018 bij de ODA's waren geregistreerd met een werkloosheidsperiode van ten minste 12 maanden (77 255 registraties).

Voor de berekening van de PSM werden verschillende variabelen in aanmerking genomen op basis van de sociaaldemografische kenmerken van de behandelde en de niet-behandelde personen, zoals: geslacht, leeftijd, staatsburgerschap, opleidingsniveau, woonplaats, datum van toetreding tot de interventie en arbeidsgeschiedenis tot 36 maanden vóór de interventie.

Voor het meten van de uitkomstvariabelen voor de behandelde groep en de controlegroep (vóór en na de interventies) werden gegevens gebruikt uit het COB, het archief waarin de verplichte kennisgeving van arbeidsovereenkomsten door bedrijven aan de ODA's is opgeslagen.

De resultaten van de studie waren gemengd: er werden positieve effecten gemeld voor stages en werkbeurzen, terwijl er negatieve effecten waren voor werkervaring in gemeenten en geen significante effecten voor opleidingsvouchers<sup>85</sup>.

### **3.2.3. Difference-in-differences**

#### *De belangrijkste kenmerken van difference-in-differences*

- Difference-in-differences is een intuïtieve methode waarbij het verschil in uitkomsten vóór en na de interventie in de behandelingsgroep wordt vergeleken met dat in de controlegroep.
- Aangezien de verandering in de loop van de tijd in de controlegroep meet wat er zonder interventie met de behandelingsgroep zou zijn gebeurd (de contrafeitelijke situatie), vertegenwoordigt elk verschil in de uitkomsten

<sup>82</sup> Zie [Caliendo, Mahlstedt en Mitnik, 2017](#) en [Kluve, Lehmann, en Schmidt, 2008](#)

<sup>83</sup> Andere voorbeelden van de praktische toepassing van methoden (PSM, RDD en DID) zijn te vinden op [Europese Commissie, 2020](#)

<sup>84</sup> Zie [Pompili M., Giorgetti I., 2020a](#)

<sup>85</sup> In het evaluatieverslag wordt onderstreept dat de bevindingen in verband met de opleidingsvouchers slechts als voorlopig kunnen worden beschouwd, aangezien de onderzochte steekproef te klein is in verhouding tot het totale aantal deelnemers.

vóór en na de interventie in de behandelingsgroep het causale effect van de interventie.

- Dit is een eenvoudige methode die in veel gevallen kan worden toegepast.

Referentiemetingen (of aan de behandeling voorafgaande metingen) van de resultaatvariabelen kunnen hetzij afzonderlijk, hetzij in combinatie met matching worden gebruikt om schattingen van difference-in-differences (DID) te verkrijgen. Hierbij wordt het verschil in een resultaat vóór en na behandeling in een controlegroep afgetrokken van hetzelfde verschil dat is waargenomen in een behandelde groep om een schatting van het effect van een interventie te verkrijgen. Ook in dit geval is de selectie van een plausibele controlegroep van fundamenteel belang. Effecten die op basis van DID worden berekend, worden gewoonlijk waargenomen binnen een regressiekader, dat ook andere waargenomen verschillen tussen de behandelings- en de controlegroep verklaart. Bovendien wordt met deze methode nagegaan of er tussen de twee groepen niet-waargenomen verschillen bestaan die onveranderd blijven in de loop van de tijd, en of er niet-waargenomen verschillen bestaan die in de loop van de tijd veranderen maar die zowel de controlegroep als de behandelingsgroep in gelijke mate beïnvloeden (bijvoorbeeld factoren die de hele economie beïnvloeden). Omdat met een DID-methode bepaalde aspecten van niet-waargenomen verschillen tussen de behandelings- en de controlegroep kunnen worden nagegaan, is deze methode in de meeste gevallen een verbetering ten opzichte van een transversale matchingstrategie. Figuur 9 geeft een visuele weergave van de DID-methode.

*Vóór en na de  
behandeling*

De x-as geeft het tijdsverloop weer en de y-as een schaal waarop de resultaten zijn geregistreerd. In dit geval zouden de resultaten lonen kunnen zijn. Het gemiddelde loon voor de behandelingsgroep in de periode voorafgaand aan de behandeling is YB1, terwijl dat voor de controlegroep YC1 is. In de periode na de behandeling bedragen de lonen respectievelijk YB2 en YC2 voor de behandelings- en de controlegroep. De ononderbroken bovenste lijn geeft dus de verandering van de lonen in de behandelingsgroep weer, terwijl de ononderbroken onderste lijn die van de controlegroep toont.

Een ruwe schatting van het effect van de interventie zou voortvloeien uit een vergelijking van de lonen in de behandelings- en de controlegroep in de periode na de behandeling, namelijk YB2 – YC2. Dit zou echter onjuist zijn, aangezien er dan geen rekening wordt gehouden met loonverschillen vóór de behandeling. Een manier om de DID-schatting te beschouwen, is die waarbij een vóór de behandeling geschatte vertekening wordt afgetrokken van het verschil in resultaten na de behandeling. Als gevolg hiervan wordt het verschil in loon na de behandeling (YB2 – YC2) gecorrigeerd door hiervan het loonverschil van vóór de behandeling (YB1 – YC1) af te trekken, zodat het effect van de DID-schatting heel eenvoudig als volgt kan worden uitgedrukt:

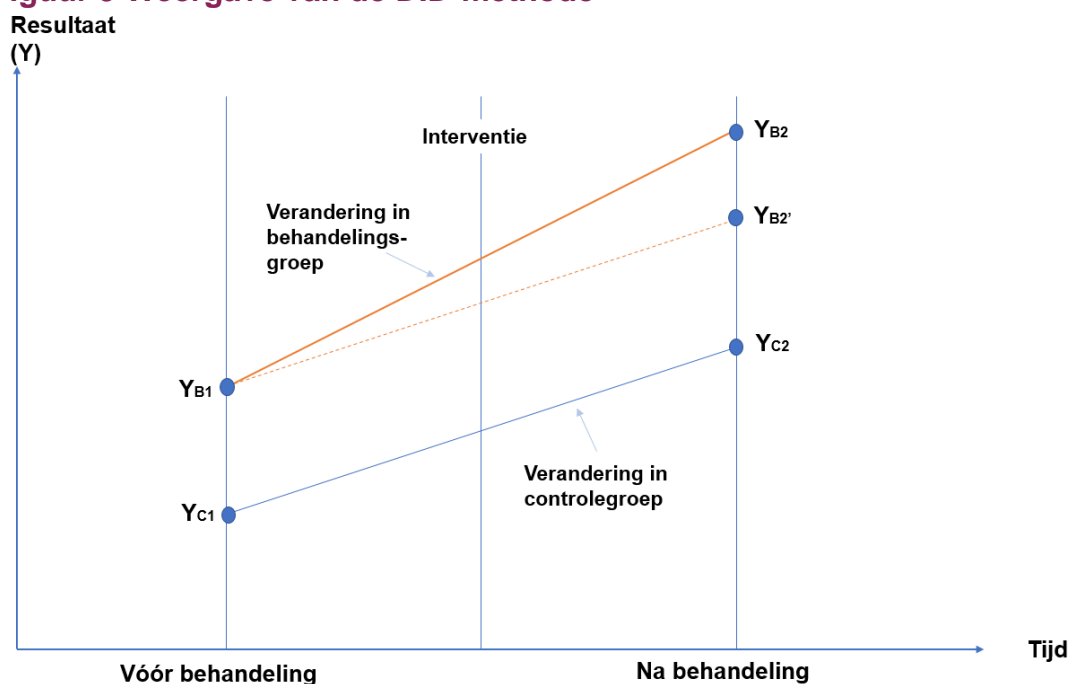
$$(YB2 - YC2) - (YB1 - YC1).$$

*De logica van het  
dubbele verschil*

Als de loonverschillen na behandeling niet worden gecorrigeerd voor eerder bestaande verschillen tussen de behandelings- en de controlegroep, kan dat tot vertekende schattingen leiden. Zoals hierboven vermeld, kan de DID-methode ook worden gezien als het aftrekken van de verandering in de resultaten voor de controlegroep van de verandering die in de behandelingsgroep is waargenomen. De waargenomen verandering in de

controlegroep wordt opgevat als dat wat zonder de interventie in de behandelingsgroep zou hebben plaatsgevonden.

**Figuur 9 Weergave van de DID-methode**



In het eenvoudigste geval is de belangrijkste aanname waarop de DID-methode berust, die van gemeenschappelijke trends: d.w.z. de tijdstrends in de resultaten binnen de behandelings- en de controlegroep zijn gelijk wanneer er geen behandeling plaatsvindt. Deze aanname kan niet rechtstreeks worden getoetst, hoewel zij de plausibiliteit van de methode ondersteunt wanneer er voor zowel de behandelings- als de controlegroep meerdere metingen van de resultaatvariabelen voorafgaand aan de behandeling beschikbaar zijn en deze metingen parallelle trends vertonen. Zie voor een voorbeeld van de DID-methode het tekstkader hieronder.

*De belangrijkste aanname: gemeenschappelijke trends zijn equivalent wanneer er geen behandeling plaatsvindt*

### Tekstkader 16 Voorbeeld van een evaluatie aan de hand van een DID-methode

#### Evaluatie van uit het ESF gefinancierde maatregelen ten behoeve van oudere werknemers in Lubelskie in de programmeringsperiode 2007-2013<sup>86</sup>

Het belangrijkste doel van de studie was de evaluatie van ESF-interventies voor de verlenging van de economische activiteit van ouderen in de ongunstige demografische context in de Poolse regio Lubelskie. De interventies werden gefinancierd in het kader van het operationele programma “Menselijk potentieel” 2007-2013.

De behandelde groep, een steekproef van 145 personen, bestond uit mensen die werkloos waren op het moment dat zij tot de interventies toetraden. De controlegroep werd verkregen op grond van geanonimiseerde registergegevens die door de arbeidsbureaus werden verstrekt en die sociaal-demografische gegevens en de geschiedenis van de registraties met betrekking tot de arbeidsstatus van specifieke personen bevatten (de gegevens waren afkomstig van acht arbeidsbureaus en bevatten informatie over 67 102 personen). De uitkomstvariabele was de “status met registratie in het werkloosheidsregister”<sup>87</sup> en deze kon zowel vóór als na de interventie worden waargenomen, zodat een DID-methode kon worden toegepast. De arbeidsstatus werd gemeten op vijf tijdstippen: 6 maanden vóór

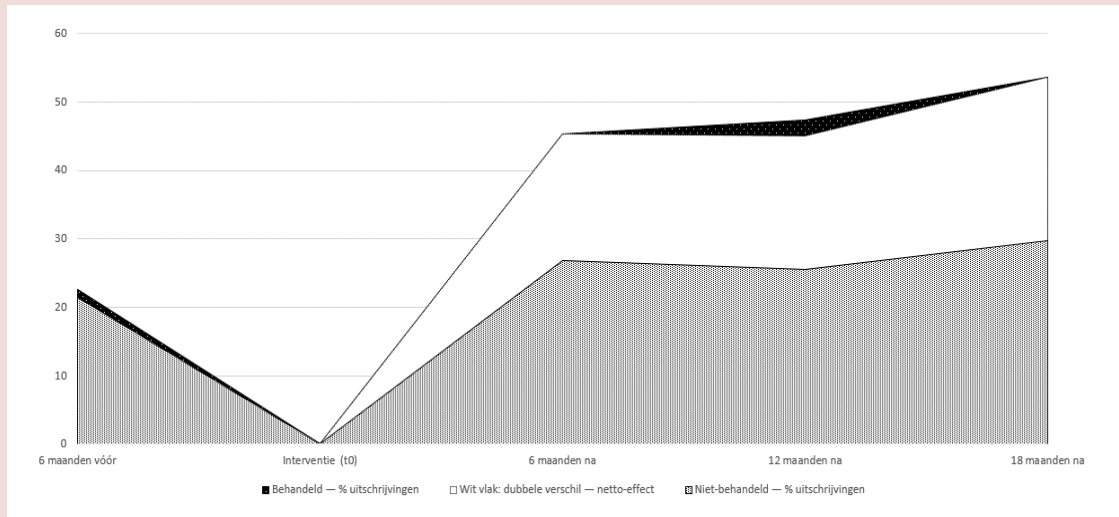
<sup>86</sup> Zie [Re-source Pracownia Badań i Doradztwa, 2015](#)

<sup>87</sup> Hoewel niet geregistreerd zijn als “werkloos” niet gelijk is aan “werkend”, was een dergelijke aanname nodig voor de uitvoering van de studie.

deelname aan het project, tijdens de deelname aan het project, en 6, 12 en 18 maanden na de beëindiging van het project.

De keuze van de controlegroep in deze difference-in-differencesmethode werd verbeterd door gebruik te maken van matchingmethoden om voor elke behandelde persoon een niet-behandelde persoon te identificeren op basis van een aantal waarneembare covarianties, zoals geslacht, leeftijd, onderwijsniveau en werkloosheidspercentage in de woonplaats.

Uit de analyse bleek dat steun voor werkloze vijftigplussers een positief netto-effect op de arbeidsparticipatie had: de behandelde werkloze vijftigplussers hadden een 2,5 keer grotere kans op uitschrijving als werkloze dan de niet-behandelde personen. Onderstaande figuur toont de ontwikkeling van de uitkomstvariabele voor de behandelde groep en de controlegroep en het geschatte netto-effect na rekening te hebben gehouden met de gelijktijdige aanname van de difference-in-differencesmethode.



### 3.2.4. Regressie-discontinuïteitsanalyse

#### Belangrijkste kenmerken van regressie-discontinuïteit

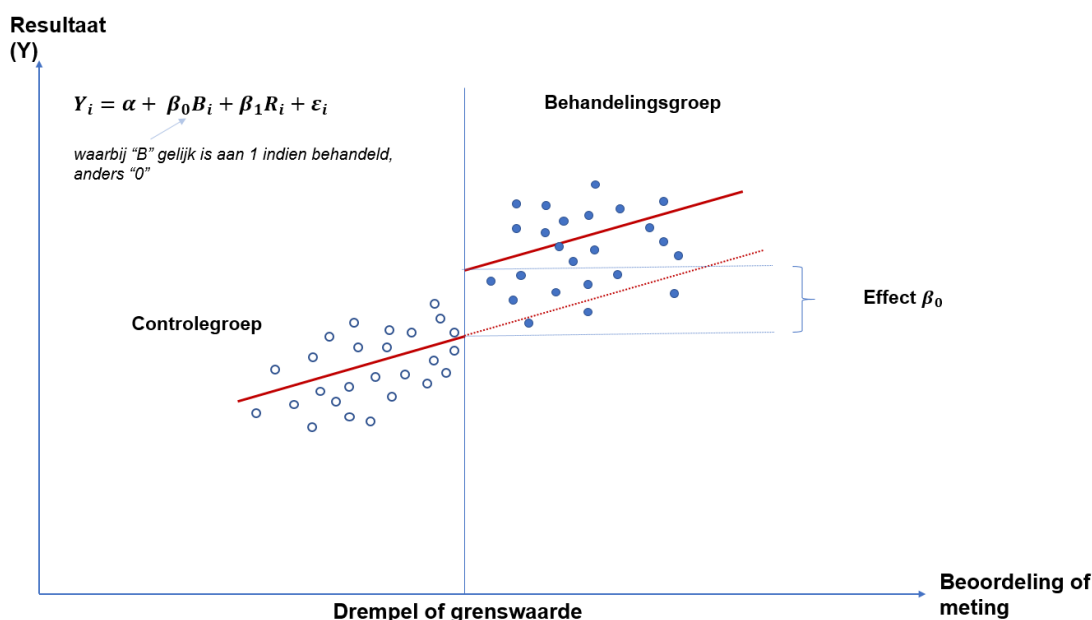
- Een regressie-discontinuïteitsanalyse is gebaseerd op het idee dat een specifieke grenswaarde ("cut-off") van een score of beoordeling bepaalt of een persoon al dan niet aan de interventie deelneemt.
- Personen die dicht bij die waarde liggen worden dan als vergelijkbaar beschouwd, met als enige verschil dat de personen aan de ene kant van de grenswaarde (de behandelingsgroep) deelnemen en die aan de andere kant (de controlegroep) niet.
- Een vergelijking van deze groepen geeft doorgaans een nauwkeurige en intuïtieve meting van het effect van de interventie; de methode kan echter alleen worden toegepast als er een selectieprocedure op basis van een score of beoordeling bestaat.

Een **regressie-discontinuïteitsmethode** kan worden toegepast wanneer de toegang tot een interventie wordt bepaald door een grenswaarde in een continue beoordeling, schaal of meting. De toegang tot opleidingen kan bijvoorbeeld worden bepaald door de prestaties in een vaardigheidstest, waarbij de personen met scores boven een bepaalde drempel (of grenswaarde) een opleiding ontvangen, en degenen die onder de drempel scoren niet. De methode is maar geldig als de onderzoeker de grenswaarde

*Een grenswaarde scheidt behandelde en niet-behandelde personen*

bepaalt zonder de scores van potentiële begunstigden van de opleiding te kennen; kandidaten met scores die dicht bij de grenswaarde liggen, zullen sterk op elkaar lijken, met het verschil dat degenen die net boven de grens liggen aan de interventie zullen deelnemen, terwijl degenen die er net onder liggen niet deelnemen. De resultaten voor personen boven en onder de grenswaarde kunnen met elkaar worden vergeleken om de effecten van de interventie bij de grenswaarde te schatten.

**Figuur 10 Weergave van de regressie-discontinuïteitsmethode**



Er kan een regressie-discontinuïteitsanalyse (*regression discontinuity design*, RDD) worden uitgevoerd wanneer de grenswaarde ofwel de behandelingsgroep volledig identificeert (met volledige overeenstemming), in welk geval een "scherpe" discontinuïteit wordt verkregen, of wanneer, onder bepaalde voorwaarden, niet alle aan een bepaalde kant van de grenswaarde gelegen personen strikt voldoen aan hun toewijzing voor behandeling (een minder scherpe of "fuzzy" discontinuïteit).

*Scherpe of fuzzy discontinuïteit*

Figuur 10 hierboven geeft een gestileerd voorbeeld van een RDD. Dit is de eenvoudigste weergave van een scherpe discontinuïteit. De interventie leidt tot constante effecten bij elke waarde van de beoordeling en de effecten worden geschat aan de hand van een lineair regressiemodel (er zijn geen problemen met betrekking tot de functionele vorm van de effectregressie). In de praktijk zal de analyse altijd veel verfijnder moeten zijn dan die in figuur 10 is weergegeven.

De stippen in figuur 10 geven individuele eenheden weer, bijvoorbeeld cursisten. De x-as geeft de beoordeling of meting weer die wordt gebruikt om cursisten toe te wijzen aan opleidingsplaatsen. Personen met een score aan de rechterkant van de verticale ononderbroken lijn die de drempel voor deze beoordeling of maatstaf aangeeft (bijvoorbeeld een vaardigheidstest), treden toe tot de opleiding en vormen de "behandelingsgroep". Potentiële cursisten die onder de drempel voor de beoordeling of meting scores, nemen niet deel aan de opleiding en vormen de controlegroep.



Het belangrijkste punt is dat de beoordeling die wordt gebruikt om de doelgroep aan behandelings- of controleomstandigheden toe te wijzen een continue kwantitatieve variabele is die vóór de behandeling wordt gemeten en dat een persoon wordt toegelaten tot het opleidingsprogramma op basis van het feit of zijn of haar score boven of onder een vooraf bepaalde grenswaarde of drempel ligt.

Het resultaat is op de y-as weergegeven. In wezen wordt het effect van de behandeling vastgesteld door op basis van de gegevens een lineair regressiemodel te schatten (in het licht van de bovenstaande aannamen); d.w.z. regressie van de resultaatvariabele ten opzichte van de beoordelingsscore samen met een dummyvariabele (een behandelingsindicator) die bepaalt of een score onder of boven de grenswaarde ligt (en dus of de persoon aan de behandelingsgroep of de controlegroep wordt toegewezen).

Een vergelijking van een dergelijke effectregressie wordt weergegeven in figuur 10. Het effect of gevolg van opleiding in dit voorbeeld wordt verkregen aan de hand van de coëfficiënt van de behandelingsindicator, namelijk  $\beta_0$ <sup>88</sup>. Dit toont op doeltreffende wijze aan of er een breuk of discontinuïteit rond de grenswaarde is, zoals aangegeven in figuur 10 door een verschuiving naar boven in de regressielijn bij de drempel of grenswaarde. In dit voorbeeld wordt een positief effect van de opleiding in het resultaat waargenomen.

Een alternatieve manier om de schatting van het effect te begrijpen, is te kijken naar de stippellijn in het verlengde van de lijn van de controlegroep, zoals weergegeven in figuur 10. Dit kan worden beschouwd als een contrafeitelijke schatting voor de behandelingsgroep — de relatie tussen de beoordeling en de meting van het resultaat die er zonder de interventie zou zijn geweest — waarbij het verschil tussen deze stippellijn en de trendlijn voor de behandelingsgroep het effect of gevolg van de behandeling aangeeft. Er wordt op gewezen dat er zonder behandeling geen sprake is van discontinuïteit in de lijn en we gaan ervan uit dat het resultaat in een dergelijk geval continu varieert naargelang de beoordeling of meting.

De regressie-discontinuïteitsmethode werkt omdat de waarnemingen in de behandelingsgroep en de controlegroep dicht bij de grenswaarde op elkaar lijken, met het verschil dat die boven de grenswaarde in dit voorbeeld een opleiding krijgen, en die eronder niet. De situatie voor waarnemingen dicht bij de grenswaarde lijkt dus op randomisatie. Er bestaat echter één grote beperking. In de meeste toepassingen kunnen de aan de hand van een RDD geschatte effecten de beleidsmaker alleen iets vertellen over effecten voor personen die dicht bij de grenswaarde of drempel liggen. De mate waarin generalisaties kunnen worden gemaakt met betrekking tot personen die verder van de drempel zijn verwijderd, kan beperkt zijn.

*Goede interne validiteit, maar generalisatie kan beperkt zijn*

RDD kan een nuttige benadering zijn waarbij personen aan een interventie worden toegewezen op basis van de behoefte die wordt gemeten aan de hand van een continue beoordeling of score. De analyse kan echter complex worden wanneer de grenswaarde onscherp is en er sprake is van niet-overeenstemming, en wanneer er problemen bestaan met betrekking tot de functionele vorm in het effectregressiemodel. Er moet van een aantal aannamen worden uitgegaan en de juistheid van deze aannamen kan niet altijd worden bevestigd.

<sup>88</sup> In een eenvoudig geval zou dit het beoogde effect zijn van een “intention to treat” bij de grenswaarde (zie [Bloom H. S., 2009](#)).

In het volgende tekstkader wordt een praktijkvoorbeeld gegeven, waarbij een RDD werd gebruikt bij de evaluatie van een jongerengarantieregeling.

### Tekstkader 17 Voorbeeld van een evaluatie waarbij een RDD wordt gebruikt

#### Beroepsopleiding voor werkloze jongeren in Letland: bewijs afkomstig van een RDD<sup>89</sup>

Onderzoekers van het Gemeenschappelijk Centrum voor onderzoek van de Europese Commissie gebruikten een RDD om het effect te beoordelen van een programma voor beroepsopleidingen op de arbeidsmarkttuitkomsten van werkloze jongeren in Letland. Het opleidingsprogramma werd gefinancierd in het kader van de jongerengarantie voor de periode 2014-2020 en was gericht op 15- tot 29-jarige NEET's.

De gebruikte gegevens waren afkomstig van het Letse staatsbureau voor arbeidsvoorziening (SEA), dat zowel informatie verstreekte over de deelnemers als over de niet-deelnemers die op specifieke data als werkloos waren geregistreerd, en werden gematched met gegevens van de belastingdienst (SRS), die informatie bevatten over het inkomen van personen op specifieke data vóór en na het programma (januari 2012 en juni 2017). Na het opschonen van de datasets telde de uiteindelijke steekproef van de behandelde groep 898 personen en die van de controlegroep 10 717 personen.

De beoordelaren baseerden zich op een regeling van de Letse regering die werklozen jonger dan 25 jaar voorrang geeft voor deelname aan het beroepsopleidingsprogramma. De leeftijd is dan ook de lopende variabele die bepalend is voor de waarschijnlijkheid van deelname aan het beroepsopleidingsprogramma, waarbij de leeftijd van 25 jaar de drempel is waaronder deelname aan het programma fors toeneemt als gevolg van de voorrangsregeling. Aangezien personen geen controle hebben over hun leeftijd, kan de toewijzing aan het opleidingsprogramma rond de drempel als willekeurig worden beschouwd. Op grond van de voorrangsregeling en het feit dat deelname vrijwillig was, moesten de onderzoekers een fuzzy RDD (FRDD) gebruiken. De identificatiestrategie was gebaseerd op het feit dat mensen dicht bij de grenswaarde (leeftijd 25 jaar) onderling vergelijkbaar waren, waarbij het enige verschil de deelname aan het beroepsopleidingsprogramma was.

Uit de bevindingen bleek dat het effect van de waarschijnlijke toekomstige arbeidsparticipatie en een maandelijks inkomen tot 3,5 jaar na toetreding tot het programma positief maar statistisch niet significant was. Er werd echter een positief effect van de voorrangsregeling ten aanzien van programmadeelname waargenomen. Aangezien de geldigheid van de FRDD berust op het feit dat potentiële deelnemers geen controle kunnen uitoefenen op de lopende variabele (leeftijd), werden er verschillende tests uitgevoerd om dit effect te bevestigen.

### 3.2.5. Instrumentele variabelen

#### Belangrijkste kenmerken van instrumentele variabelen

- Het idee van instrumentele variabelen is dat een bepaalde (vooraf bepaalde/exogene) procedure de deelname bepaalt, maar dat dit niet de feitelijke selectieprocedure is.
- De vooraf bepaalde procedure en de belangrijkste variabelen ervan kunnen worden gebruikt om rekening te houden met zelfselectie in de beslissing van personen om al dan niet deel te nemen.

Bij de toepassing van de **instrumentele-variabelen**methode (IV) moet de selectie voor deelname aan de behandeling ten minste gedeeltelijk worden bepaald door een exogene factor die geen verband houdt met andere resultaten dan die via de behandeling worden verkregen. Als zodanig is de exogene factor van invloed op de deelname, maar niet rechtstreeks op de resultaten. Doorgaans kunnen dergelijke exogene factoren administratieve

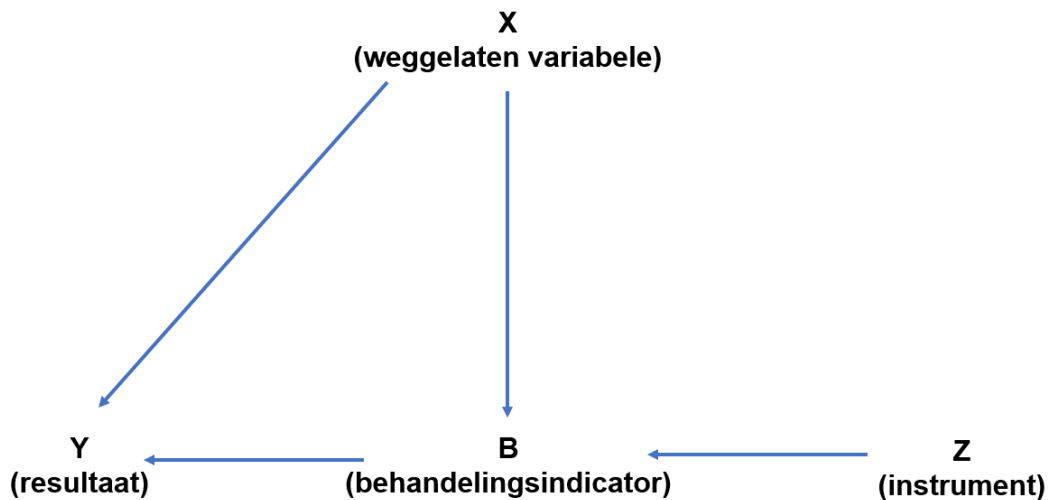
*Hoe exogene factoren de deelname beïnvloeden*

<sup>89</sup> Zie [Bratti M. et al., 2018](#)

fouten of vergissingen zijn, of andere toevallige verschillen in de ontvangen behandeling.

Figuur 11 hieronder geeft de instrumentele-variabelenmethode weer. Er worden vier variabelen in een zeer vereenvoudigd causaal systeem getoond. De variabelen zijn de gegevens die zijn verzameld van een populatie die hypothetisch gezien de doelgroep is van een opleidingsinterventie (zowel degenen die een opleiding krijgen als de personen in de controlegroep).

**Figuur 11 Weergave van een instrumentele-variabelenmethode**



“Y” staat voor het onderzochte resultaat. In geval van een opleidingsinterventie kan dat bijvoorbeeld het loon zijn. “B” is een indicator die aangeeft of een persoon is toegetreden tot het opleidingsprogramma<sup>90</sup>.

“X” is een weggelaten variabele (“omitted variable”) die niet wordt waargenomen, maar die verband houdt met zowel Y (het resultaat) als B (de behandelingsindicator); in het geval van een opleidingsprogramma zou dit bijvoorbeeld een referentiemeting van de vaardigheid kunnen zijn. In dit geval houdt vaardigheid verband met zowel deelname aan de opleiding als inkomen. Zo kunnen vaardigere leden van de doelgroep ervoor kiezen om een opleiding te volgen en een hoger loon ontvangen.

Het bestaan van X rechtvaardigt het zoeken naar een instrument en betekent dat het effect van de opleiding op het inkomen wordt beïnvloed door dit bestaan van X. Met andere woorden, de schatting is vertekend vanwege het bestaan van X en het feit dat deze variabele niet wordt waargenomen en niet rechtstreeks in de analyse kan worden verwerkt.

Tot slot is de variabele “Z” een instrument dat volgens Morgan en Winship<sup>91</sup> kan worden beschouwd als een schok die op B inwerkt en die onafhankelijk is van X. Om deze reden is er in figuur 11 geen lijn die Z met X verbindt. Bovendien heeft Z alleen invloed op Y via B, en is er geen andere weg waarlangs Z invloed heeft op Y. Dit betekent dat Z kan worden gebruikt om in B (de behandeling) variatie aan te brengen die geen verband houdt met de

*De identificatie van een geldig instrument is van fundamenteel belang*

<sup>90</sup> Met andere woorden, er is volledige overeenstemming en alle personen in de behandelingsgroep nemen deel aan de opleiding.

<sup>91</sup> Morgan S.L., Winship C., 2014.

verstorende variabele X. Als gevolg hiervan kan een zuivere meting van het effect van B op Y worden verkregen door deze variatie te gebruiken<sup>92</sup>.

De eenvoudigste omstandigheden waarin een IV-methode zou kunnen worden gevolgd, worden hieronder beschreven, waarbij een groot aantal complexe factoren buiten beschouwing worden gelaten. In de praktijk is het vaak moeilijk om een geschikt instrument te vinden. De plausibiliteit van verschillende potentiële instrumenten is sterk contextgebonden en de onderliggende aannamen voor identificatie kunnen over het algemeen niet statistisch worden getoetst (in werkelijkheid kan de vereiste correlatie tussen Z en B statistisch worden getoetst, terwijl dat voor de “onafhankelijkheid” van Z en X niet mogelijk is). Een mogelijke strategie is om de afstand van de opleidingscentra (de fysieke locatie van de opleiding) tot de woonplaats van een cursist te gebruiken als instrument voor het schatten van het effect van opleiding op het nettoloon van de cursisten. Mogelijk wordt waargenomen dat cursisten die dichter bij opleidingscentra wonen, eerder aan een opleidingsinterventie deelnemen en dat de afstand tussen de woonplaats van een cursist en een opleidingscentrum geen verband houdt met andere factoren die bepalend zijn voor nettoloon en deelname aan de opleiding (bijvoorbeeld maatregelen inzake menselijk potentieel). De enige wijze waardoor deze afstandsmeting van invloed kan zijn op het nettoloon, is via het effect ervan op de opleiding<sup>93</sup>.

Instrumentele variabelen kunnen in een uitgebreide reeks situaties worden gebruikt. Afhankelijk van de responsvariabele, kunnen schattingen worden gemaakt aan de hand van verschillende schattingsmethoden. Tot dusver is deze methode niet vaak gebruikt in ESF-evaluaties. In tekstkader 18 hieronder tonen we een voorbeeld van de analyse van de causale effecten van een beleid met gebruik van een IV-methode.

---

<sup>92</sup> Het causale effect van B op Y wordt berekend in aanwezigheid van een instrument door de verhouding tussen Z en Y te schatten, en deze schatting te delen door de geschatte verhouding tussen Z en B.

<sup>93</sup> De interpretatie van bevindingen van een dergelijke analyse kan gecompliceerder worden door de vraag of het instrument al dan niet verband houdt met de variatie in het effect van de behandeling (zie [Bryson, et al., 2002](#)).

**Tekstkader 18 Voorbeeld van een studie met gebruik van een instrumentele-variabelenmethode****Effecten van een taal cursus voor werkloze immigranten op de arbeidsparticipatie<sup>94</sup>**

Het verwerven van taalvaardigheid in de taal van het gastland is belangrijk voor immigranten om een hoog niveau van integratie op de arbeidsmarkt te bereiken.

In een recent document (Lang J., 2021) gebruikt de auteur een IV-methode om een schatting te maken van het causale effect, twee jaar na inschrijving van de deelnemers, van een cursus Duits voor beroepsmatige doeleinden op de arbeidsparticipatie van immigranten. Het programma werd uitgevoerd door het Duitse federale bureau voor migratie en vluchtelingen (BAMF) en werd gefinancierd uit het ESF; het ESF-BAMF-programma liep tot eind 2017. De cursussen omvatten verschillende onderdelen: een cursus Duits, ontwikkeling van de beroepsvaardigheden en stages. Mensen met een migratieachtergrond die actief waren op de Duitse arbeidsmarkt kwamen voor het programma in aanmerking, ongeacht hun nationaliteit en datum van immigratie.

Om de niet-waargenomen heterogeniteit op het gebied van taalvaardigheden aan te pakken, maakt de auteur gebruik van de exogene variatie in intensiteit van de lokale taal cursussen bij de arbeidsbureaus. Dit wordt beschouwd als een geldig instrument, aangezien arbeidsbureaus een ruime beoordelingsvrijheid hebben bij de uitvoering van verschillende combinaties van programma's, afhankelijk van hun "beleidsstijl", en deze variabele is exogeen ten aanzien van de uitkomsten van de arbeidsparticipatie onder de werkzoekenden. De analyse betrof cursussen die in 2014 werden uitgevoerd.

De evaluatie is gebaseerd op een uitgebreide reeks administratieve gegevens, geïntegreerde beroepsbiografieën (IEB), en een samengevoegde databank van administratieve informatie van het Duitse federale arbeidsbureau. De IEB's bevatten informatie over deelname aan taal cursussen, gedetailleerde informatie over het arbeidsverleden (met uitzondering van zelfstandige arbeid), het zoeken naar werk, uitkeringen tijdens de werkloosheidsperiode en persoonsgegevens. Er werden ook andere gegevens (WGH, Werdegangshistorie) gebruikt om gedetailleerde informatie te verkrijgen over de deelname aan de door BAMF verstrekte taal cursussen en om, waar mogelijk, in de IEB ontbrekende gegevens aan te vullen met informatie over perioden van zelfstandige arbeid en ouderschapsverlof.

Uit de bevindingen is gebleken dat na een periode van lock-in-effecten van iets meer dan een jaar, de immigranten die deelnemen aan het opleidingsprogramma twee jaar na de interventie een kans hebben om in dienst te worden genomen, die negen procentpunten hoger is dan die van de controlegroep, en dat deze kans niet beperkt is tot ongeschoold werk. Ook werd geconstateerd dat het vroegtijdig aanbieden van taalopleidingen (kort na aankomst in Duitsland) de integratie van immigranten op de arbeidsmarkt ten goede komt.

---

<sup>94</sup> Zie [Lang, J., 2021](#)

**Tabel 5 Vergelijking van de hoofdkenmerken van de belangrijkste CIE-methoden**

Methode	Belangrijkste kenmerken	Voordelen	Gegevensvereisten	Beperkingen
Randomisatie — experimentele methode	<ul style="list-style-type: none"> <li>- Personen die voor deelname in aanmerking komen, worden willekeurig toegewezen aan een behandelings- en een controlegroep.</li> <li>- Randomisatie zorgt ervoor dat beide groepen (in het algemeen) identiek zijn wat alle relevante kenmerken betreft.</li> <li>- De controlegroep geeft derhalve het antwoord op de contrafeitelijke vraag, en het verschil in de uitkomsten tussen de behandelingsgroep en de controlegroep geeft het causale effect van de interventie aan.</li> </ul>	<ul style="list-style-type: none"> <li>- Mits goed uitgevoerd, zijn de effectschattingen “zuiver”.</li> <li>- De resultaten zijn transparant en eenvoudig te begrijpen.</li> <li>- De bevindingen zijn minder onderhevig aan voorbehoud en twijfel.</li> <li>- Verscheidenheid aan beschikbare opzetvarianten om een reeks beleidscontexten en interventieomstandigheden te kunnen aanpakken.</li> </ul>	<ul style="list-style-type: none"> <li>- Basisvereiste om de selectie voor de interventie via randomisatie uit te voeren.</li> <li>- Registratie van wie is toegewezen aan welke groep.</li> <li>- Verzameling van referentiegegevens is essentieel.</li> <li>- Resultaatmetingen moeten worden vastgelegd voor zowel de behandelings- als de controlegroep.</li> </ul>	<ul style="list-style-type: none"> <li>- Vaak voldoen de behandelings- en de controlegroep niet aan de toewijzingsregels voor de behandeling.</li> <li>- Vaak is toestemming van de deelnemers vereist.</li> <li>- Randomisatie kan de selectie beïnvloeden van degenen die deelnemen aan/een aanvraag indienen voor een interventie.</li> <li>- Het feit dat deelnemers zich bewust zijn van hun toewijzingsstatus kan hun gedrag veranderen en de resultaten beïnvloeden.</li> <li>- Ethische overwegingen.</li> <li>- Aanzienlijke vereisten wat betreft planning en opzet.</li> <li>- Kan duur zijn (maar is dat niet noodzakelijkerwijs).</li> </ul>
Matching (propensity-score)	<ul style="list-style-type: none"> <li>- Net als bij randomisatie wordt bij matching achteraf een controlegroep samengesteld die voor alle relevante kenmerken zo vergelijkbaar mogelijk is (in het algemeen) met de behandelingsgroep.</li> <li>- Anders dan bij randomisatie kunnen alleen waarneembare kenmerken (leeftijd, geslacht, onderwijs enz.) worden vergeleken, terwijl niet-waarneembare kenmerken (zoals motivatie) niet in aanmerking kunnen worden genomen.</li> <li>- De geldigheid van de methode hangt dus sterk af van de beschikbaarheid van gegevens.</li> </ul>	<ul style="list-style-type: none"> <li>- Vereist een gedegen kennis van selectieprocedures, maar geen rechtstreekse controle over de selectie voor de interventie.</li> <li>- Kan met terugwerkende kracht worden toegepast, indien de juiste gegevens beschikbaar zijn, en in verschillende contexten.</li> <li>- Technisch gezien een semi-parametrische schattingsmethode; vereist minder parametrische aannamen (bijvoorbeeld geen behoefte aan een standaardaanname voor regressie).</li> <li>- Kan worden gebruikt om de effecten van meervoudige behandelingen te</li> </ul>	<ul style="list-style-type: none"> <li>- Nauwkeurige identificatie van de deelnemers aan de interventie.</li> <li>- Gegevensbronnen waaruit de steekproef kan worden getrokken.</li> <li>- Duidelijk concept van deelname en goed begrip van selectie voor behandeling.</li> <li>- Uitgebreide gegevens, het liefst verzameld in de uitgangssituatie met het oog op matching.</li> <li>- Resultaatmetingen van de interventie voor deelnemers en controle-eenheden.</li> </ul>	<ul style="list-style-type: none"> <li>- Vereist aanzienlijke hoeveelheden gegevens die een volledige karakterisering van de selectieprocedure mogelijk maken.</li> <li>- De geldigheid hangt af van de kwaliteit van de controlegroep en de zorgvuldige selectie ervan, en van de mate van common support.</li> <li>- Berust op de aanname dat de selectie voor de interventie op passende wijze kan worden gekenmerkt door waarneembare gegevens.</li> <li>- De verschillende beschikbare matchingmethoden vereisen een gevoelige analyse.</li> </ul>

Methode	Belangrijkste kenmerken	Voordelen	Gegevensvereisten	Beperkingen
		<p>schatten.</p>		<ul style="list-style-type: none"> <li>- De resultaten kunnen complex zijn om uit te leggen en te interpreteren, en zijn mogelijk ambigu.</li> </ul>
<p>Difference-in-differences (DID)</p>	<ul style="list-style-type: none"> <li>- DID is een intuïtieve methode waarbij het verschil in uitkomsten vóór en na de interventie in de behandelingsgroep wordt vergeleken met het verschil vóór en na de interventie in de controlegroep.</li> <li>- Aangezien de verandering in de loop van de tijd in de controlegroep meet wat er zonder de interventie met de behandelingsgroep zou zijn gebeurd (de contrafeitelijke situatie), vertegenwoordigt elk verschil in de uitkomsten vóór en na de interventie in de behandelingsgroep het causale effect van de interventie.</li> <li>- Dit is een eenvoudige methode die in veel gevallen in de praktijk kan worden toegepast.</li> </ul>	<ul style="list-style-type: none"> <li>- Controleert bepaalde aspecten van niet-waargenomen verschillen tussen deelnemers en controlegroep.</li> <li>- Kan worden gebruikt in combinatie met matching.</li> <li>- Werkt met gegevens van vóór en na interventie, zoals panelgegevens (individuele gegevens in de tijd) of herhaalde transversale gegevens (gegevens over personen die op verschillende tijdstippen zijn verzameld).</li> </ul>	<ul style="list-style-type: none"> <li>- De gegevensvereisten zijn vergelijkbaar met die van andere benaderingen, maar met de aanvullende eis van resultaatmetingen vóór interventie.</li> <li>- Om de belangrijkste aannamen te toetsen, zijn er meerdere waarnemingen van de resultaten vóór de behandeling nodig voor zowel de behandelings- als de controlegroep.</li> </ul>	<ul style="list-style-type: none"> <li>- Vereist de aanname van gemeenschappelijke trends in de resultaten tussen de deelnemers en de controle-eenheden.</li> <li>- De analyse kan vrij complex worden en verkeerd worden geïnterpreteerd.</li> <li>- Er zijn uitgebreide gegevens over de resultaten vóór de behandeling nodig om de aanname van gemeenschappelijke trends te toetsen.</li> <li>- Kan niet worden gebruikt om de effecten van meervoudige behandelingen te schatten<sup>95</sup>.</li> </ul>
<p>Regressie-discontinuïteitsanalyse (RDD)</p>	<ul style="list-style-type: none"> <li>- Een RDD is gebaseerd op het idee dat een specifieke grenswaarde ("cut-off") van een score of beoordeling bepaalt of een persoon al dan niet aan de interventie deelneemt.</li> <li>- Personen die dicht bij die waarde liggen worden dan als vergelijkbaar beschouwd, met als enige verschil dat de personen aan de ene kant van de grenswaarde (de behandelingsgroep) deelnemen en die aan de andere kant (de controlegroep) niet.</li> <li>- Een vergelijking van deze groepen geeft doorgaans een nauwkeurige en intuïtieve meting van het effect van de interventie; de methode kan echter alleen worden toegepast als er een selectieprocedure op basis van een score of beoordeling bestaat.</li> </ul>	<ul style="list-style-type: none"> <li>- Er zijn zowel scherpe als onscherpe ("fuzzy") benaderingen van RDD beschikbaar.</li> <li>- Kan onder bepaalde voorwaarden zuivere effecten van een behandeling bieden.</li> </ul>	<ul style="list-style-type: none"> <li>- De keuze van de grenswaarde moet onafhankelijk zijn van de waarden van de beoordeling die aan elk lid van een doelgroep worden gegeven.</li> <li>- Gegevens over personen zijn vereist voor zowel de behandelings- als de controlegroep wat betreft de beoordeling of meting, de drempel of de grenswaarde en de resultaten.</li> </ul>	<ul style="list-style-type: none"> <li>- Deze benadering is niet geldig zonder een continue meting of beoordeling die bepalend is voor de behandeling.</li> <li>- Analyses kunnen complex en onzeker worden wanneer zich problemen in verband met de functionele vorm van de effectregressie voordoen, wanneer er sprake is van niet-overeenstemming of wanneer de omvang van de steekproef rond de grenswaarde beperkt is.</li> <li>- Er kunnen problemen ontstaan bij het interpreteren van bevindingen en het afleiden van generalisaties.</li> </ul>

<sup>95</sup> Zie [Frolich, M., 2004](#)

Methode	Belangrijkste kenmerken	Voordelen	Gegevensvereisten	Beperkingen
Instrumentele variabelen (IV)	<ul style="list-style-type: none"> <li>- Het idee van IV is dat een bepaalde (vooraf bepaalde/exogene) procedure de deelname bepaalt, maar dat dit niet de feitelijke selectieprocedure is.</li> <li>- Deze vooraf bepaalde procedure kan vervolgens worden gebruikt om rekening te houden met elke zelfselectie in de feitelijke van personen om al dan niet deel te nemen.</li> </ul>	<ul style="list-style-type: none"> <li>- Kan hoogwaardige schattingen of bewijs van het bestaan van causale effecten opleveren.</li> <li>- Lost het probleem van vertekening vanwege weggelaten variabelen (of selectievertekening) op.</li> <li>- Kan met terugwerkende kracht worden gebruikt.</li> </ul>	<ul style="list-style-type: none"> <li>- Vereist referentiegegevens, gegevens over resultaten en deelname aan de interventie, en ook dat er een instrument kan worden geïdentificeerd.</li> <li>- Een instrument moet verband houden met de deelname aan de interventie en heeft alleen zo gevolgen voor de resultaten. Het instrument mag geen verband houden met andere determinanten van de resultaten.</li> </ul>	<ul style="list-style-type: none"> <li>- Het kan problematisch zijn een plausibel instrument te vinden.</li> <li>- Kan moeilijk uit te leggen zijn aan niet-deskundigen.</li> <li>- De interpretatie van de resultaten is niet eenvoudig; beperkte toetsbaarheid van de vaststelling van aannamen.</li> </ul>



# Hoofdstuk 4. Hoe nu verder?

## Een agenda voor CIE's

Deze leidraad is bedoeld om beheerautoriteiten aan te moedigen en te ondersteunen bij het uitvoeren van meer wijdverbreide en kwalitatief hoogwaardige CIE's. Hij verschaft richtsnoeren voor degenen die verantwoordelijk zijn voor de planning en het laten uitvoeren van effectonderzoeken betreffende uit het ESF+ medegefinancierde interventies. Tot dusver is de nadruk gelegd op de planning en uitvoering van een CIE en zijn een aantal belangrijke vraagstukken besproken die aandacht behoeven. In de programmeringsperiode 2014-2020 werd echter gewezen op een aantal andere "bredere kwesties" en op uitdagingen die in aanmerking moeten worden genomen om de ESF+-evaluaties en het gebruik van CIE's te versterken.

In dit deel van de leidraad worden enkele suggesties gedaan om deze "bredere" kwesties aan te pakken. In het bijzonder worden de te nemen stappen in verband met de volgende punten besproken:

*Aan te pakken bredere kwesties*

- een gebrek aan kennis van CIE-methoden binnen beheerautoriteiten en in de bredere beleidsvormingsgemeenschap van de lidstaten;
- een gebrek aan externe, naar behoren gekwalificeerde en ervaren contractanten in de lidstaten die in staat zijn om CIE's uit te voeren;
- het aanpakken van juridische belemmeringen die in het algemeen bij CIE's voorkomen;
- streven naar een betere planning van CIE's;
- uitbreiding van het toepassingsgebied van CIE's.

### 4.1. Verbetering van het inzicht bij belanghebbenden

Voor de programmeringsperiode 2021-2027 is in artikel 44, lid 1, van de GB-verordening bepaald: "[de beheerautoriteit] verricht evaluaties van de programma's met betrekking tot een of meer van de volgende criteria: effectiviteit, efficiëntie, relevantie, samenhang en Uniemeerwaarde, dit ter verbetering van de kwaliteit van het ontwerp en de uitvoering van de programma's. De evaluaties kunnen ook betrekking hebben op andere relevante criteria, zoals inclusiviteit, non-discriminatie en zichtbaarheid, en op meer dan één programma". Zoals in de vorige punten is besproken, betekent dit dat er voldoende evaluatiecapaciteit moet bestaan bij zowel de beheerautoriteit die het CIE uitbesteedt als bij de beoordelaren die op het contract inschrijven.

In sommige gevallen registreerden de diensten van de evaluatiehelpdesk bij bepaalde overheidsdiensten een gebrek aan capaciteit om CIE's uit te voeren, namelijk bij de aanbesteding en de begeleiding bij de uitvoering. Dit maakte het voor de beoordelaren moeilijk om CIE's uit te voeren omdat vooraf geen duidelijke evaluatievragen waren geformuleerd, de beschikbaarheid van basisgegevens niet was vastgesteld en geen goed geïnformeerde planning van de uitvoering was opgemaakt.

Soms, vooral in kleine landen, ontbreekt op de consultancymarkt de capaciteit om een CIE uit te voeren, evenals een aanbod van technische expertise. Daarom bestaat er een algemene behoefte om zowel de vraag naar als het aanbod van CIE's te stimuleren. Wat het aanbod betreft, kan de situatie verbeteren nu de beheerautoriteiten en lidstaten opdrachten voor CIE's beginnen te geven of hun vraag naar dergelijke studies bekendmaken. De snelheid waarmee kan worden gereageerd op de toegenomen vraag naar CIE's zal afhangen van de bestaande vaardigheden en ervaring en van de aanwezigheid van instellingen in de lidstaten die in staat zijn dergelijke benaderingen toe te passen. Het stimuleren van de vraag kan echter ook gedeeltelijk worden bereikt door de kennis van en het inzicht in CIE-methoden te verbeteren bij degenen die bij beheerautoriteiten werken.

*De vraag naar en het aanbod van CIE's stimuleren*

Een oplossing voor dit probleem is dat beheerautoriteiten opleidingen over CIE-methoden voor hun personeel organiseren. Die opleidingen zouden gericht moeten zijn op de voordelen die de invoering van CIE-methoden voor beheerautoriteiten met zich meebrengen. Daarnaast moet de nadruk worden gelegd op het afleggen van verantwoording en op het leren van wat werkt. Bijlage 2 bevat een voorstel voor de opzet van een cursus.

*Ontwikkeling van opleiding op het gebied van CIE-methoden*

## 4.2. Capaciteitsontwikkeling

Een ander punt dat in de periode 2014-2020 aan de orde werd gesteld en dat in de bovenstaande delen is genoemd, was dat in kringen van onderzoek en wetenschap en bij adviesbureaus in de lidstaten de capaciteit moet worden ontwikkeld om CIE's uit te voeren. In sommige gevallen bleek dat de lidstaten over de vereiste vaardigheden beschikten, maar dat de personen met deze vaardigheden problemen hadden ondervonden om ze in een evaluatiecontext toe te passen (bv. beperkte toegang tot bruikbare gegevens of problemen bij de samenstelling van een goede controlegroep).

Er kunnen meerdere stappen worden genomen om de verstrekking van evaluatiediensten te ontwikkelen. Veel van de aan de orde gestelde kwesties gelden evenzeer voor CIE's als voor evaluaties in het algemeen. Er worden normaal gesproken drie stappen ondernomen om het aanbod op het gebied van evaluatie te verbeteren:

*Versterking van instellingen en opbouw van praktijkgemeenschappen*

- relaties opbouwen met academische instellingen, met name universiteiten;
- een onafhankelijke gemeenschap van consultants ontwikkelen en versterken;
- de ontwikkeling van een professionele evaluatiegemeenschap ondersteunen.

### Universiteiten

Het ontwikkelen van banden met universiteiten is om twee redenen belangrijk. Ten eerste beschikt academisch personeel aan universiteiten mogelijk over de vaardigheden en kennis die nodig zijn om CIE's uit te voeren. Veel micro-economen, econometristen, sociologen of psychologen met een specialisatie in kwantitatief onderzoek beschikken bijvoorbeeld over het soort vaardigheden dat is vereist. In veel lidstaten

*Ontwikkeling van academische vaardigheden*

kunnen de vereiste vaardigheden aanwezig zijn, maar zijn zij door degenen die erover beschikken tot dusver niet toegepast bij de evaluatie van interventies. Het kan zijn dat deze personen niet gestimuleerd zijn om dit te doen, wat een punt is dat moet worden aangepakt.

In sommige lidstaten bestaat de traditie om academische onderzoekers actief te betrekken bij toegepast beleidsonderzoek. In dergelijke gevallen zullen academici gewend zijn om met de regering en beheerautoriteiten samen te werken. In andere lidstaten, waar universiteiten en academici niet actief zijn op het gebied van toegepast onderzoek, kan een cultuuromslag nodig zijn. Een succesvolle methode voor het ontwikkelen van een leveranciersbasis binnen de universitaire sector is het financieren van de kosten van onderzoekscentra voor CIE-methoden door de autoriteiten van de lidstaten en de beheerautoriteiten.

Universiteiten en academici kunnen ook een rol spelen bij het opleiden van de volgende generatie beoordelaren. Wanneer zij nauw samenwerken met universiteiten, kunnen zij worden aangemoedigd om in hun curricula evaluatiemethoden op te nemen en er in het kader van deze ontwikkeling voor te zorgen dat CIE-methoden in onderwijsprogramma's worden opgenomen. In sommige lidstaten kunnen universiteiten ook een rol spelen bij het organiseren van permanente bijscholing op het gebied van effectonderzoek en CIE-methoden. Dergelijke opleidingen kunnen gericht zijn op beleidsmakers, technische specialisten binnen beheerautoriteiten en andere potentiële leveranciers zoals onafhankelijke consultants. De lidstaten kunnen overwegen financiering voor deze opleidingen te verstrekken.

*De volgende generatie opleiden*

### **Onafhankelijke consultants**

Voor sommige grootschalige evaluaties is er een internationale markt. Dit is zeker het geval voor grote CIE's. Veel lidstaten willen echter binnenlandse capaciteit ontwikkelen om CIE's uit te voeren. Een strategie om dit te bereiken kan bestaan in het tot stand brengen van strategische allianties tussen potentiële binnenlandse leveranciers en internationale consultancybedrijven.

*De markt ontwikkelen*

Hieronder volgen verschillende suggesties voor de ontwikkeling van een binnenlandse leveranciersbasis om CIE's uit te voeren, die kunnen worden toegepast door beheerautoriteiten (of andere instanties) die CIE's in opdracht geven:

- erop aandringen dat inschrijvingen van consortia of partnerschappen altijd een aantal lokale consultants omvatten;
- het volume van evaluatiecontracten dusdanig bepalen dat relatief kleine, weinig risicovolle evaluaties kunnen worden uitgevoerd door nationale nieuwkomers op de evaluatiemarkt;
- ervoor zorgen dat de technische en financiële vereisten in verband met inschrijvingen voor evaluaties niet te beperkend zijn en de deelname van nieuwkomers mogelijk maken;
- de nadruk leggen op technische criteria en knowhow tijdens de selectie in plaats van op complexe administratieve procedures waarmee minder ervaren consultants mogelijk niet vertrouwd zijn;

- het houden van briefings met potentiële consultants om vragen te beantwoorden en inschrijvingen in een concurrerende omgeving aan te moedigen;
- netwerkvorming tussen relatief geïsoleerde evaluatieconsultants ondersteunen, om teamvorming, de vorming van consortia en andere professionele netwerken en verenigingen, ook op internationaal niveau, aan te moedigen;
- de opdrachtgevers van een evaluatie doen erkennen dat zij wellicht een meer praktijkgerichte beheersbenadering voor nieuwe contractanten moeten volgen zodat deze sneller kennis en ervaring kunnen opdoen.

### Tekstkader 19 Voorbeeld van een project ter versterking van de CIE-cultuur en -capaciteit

In de programmeringsperiode 2014-2020 werd in Zweden een ESF-project uitgevoerd met als doel de cultuur en de evaluatiecapaciteit van de openbare diensten voor arbeidsvoorziening (ODA's) te verbeteren. Het project, genaamd "door de EU gefinancierde empirisch onderbouwde projecten", liep van 2016 tot 2018 en betrof managers van ODA's en ander personeel bij workshops en andere opleidingen op het gebied van evaluatie.

De activiteiten waren niet gericht op technische kwesties, maar op het belang van een evaluatie, de belangrijkste stappen van een effectonderzoek, het belang van goede gegevens en de kwaliteitsborging van een evaluatie. Er werd een website opgezet waarop educatief materiaal werd geüpload.

De belangrijkste resultaten van het project waren de verbetering van de kennis van de ODA-managers op het gebied van evaluaties, hun bereidheid evaluaties uit te voeren en hun coöperatieve houding. Concreet zijn volgens de geïnterviewde persoon "de effecten van de projecten op het aantal uitgevoerde effectonderzoeken" nog niet groot, maar is er enige vooruitgang waar te nemen, aangezien er in de vorige programmeringsperiode 2007-2013 in Zweden geen CIE's zijn uitgevoerd.

De evaluatie van het project "Ung framtid" ("Jonge toekomst"), waarin voor een gerandomiseerde benadering was gekozen, is een concreet voorbeeld in de praktijk, waarin ODA-managers betrokken waren bij het evaluatieproces<sup>96</sup>.

### Professionele gemeenschap

Het is belangrijk om professionele evaluatiegemeenschappen in de lidstaten te ontwikkelen, waar ruimte is voor discussie over CIE-methoden en het delen van ervaringen. De ontwikkeling van dergelijke gemeenschappen draagt bij tot wederzijdse ondersteuning en wederzijds leren en tot het handhaven van kwaliteitsnormen. Een nuttige strategie zou kunnen zijn contacten tot stand te brengen met de betrokken nationale evaluatieorganisaties en hen aan te moedigen CIE's van ESF+-interventies te promoten tijdens opleidingsevenementen, specifieke conferenties of seminars en bewustmakingssessies.

*Professionele gemeenschappen ontwikkelen*

### Ervaringen delen

De Commissie wil graag dat er rigoureuzere ESF+-effectonderzoeken worden uitgevoerd, en in de periode 2014-2020 werden CIE's ten sterkste aanbevolen<sup>97</sup>. Op dit moment kan worden gezegd dat het gebruik van CIE's

*Bestaande fora benutten*

<sup>96</sup> De informatie is afkomstig van een interview met een ambtenaar van de Zweedse ODA Arbetsförmedlingen.

<sup>97</sup> In de verordening gemeenschappelijke bepalingen voor de periode 2014-2020 (Verordening 1303/2013), bijlage XI, werd gevraagd om een "doeltreffend systeem van resultaatindicatoren" om "toezicht te houden op de voortgang op het gebied van resultaten en om een effectbeoordeling uit te voeren". Dit vereiste maakte deel uit van de ex-antevoorwaarden voor de

is toegenomen, maar dat zij nog steeds beperkt zijn in een aantal beheerautoriteiten en lidstaten. Het delen van ervaringen met CIE-methoden is een van de belangrijkste manieren om capaciteiten te ontwikkelen en de verspreiding van CIE's in de hele EU te ondersteunen. Daartoe moet gebruik worden gemaakt van bestaande fora voor wederzijds leren, zoals peer reviews met betrekking tot beleidsmaatregelen op het gebied van werkgelegenheid en sociale inclusie, en van praktijkgemeenschappen binnen het ESF+. Ook initiatieven van individuele beheerautoriteiten of lidstaten, zoals internationale conferenties of seminars, kunnen de uitwisseling van ervaringen bevorderen.

### 4.3. Juridische belemmeringen aanpakken

Een van de belangrijkste en grootste problemen voor onderzoekers die in de lidstaten CIE's uitvoeren, is het verkrijgen van toegang tot gegevens. Met name stuiten onderzoekers vaak op juridische belemmeringen die tot doel hebben de vertrouwelijkheid te beschermen van de gegevens van de personen waarop de datasets betrekking hebben. Zoals hierboven met betrekking tot de kenmerken van de AVG is getoond, ligt het antwoord op deze problemen in het doorvoeren van bredere hervormingen en het sluiten van overeenkomsten die het mogelijk maken relevante gegevens op een gecontroleerde manier en op permanente basis ter beschikking van beoordelaren te stellen.

*Juridische belemmeringen voor toegang tot gegevens aanpakken*

Analytische versies van administratieve datasets kunnen op regelmatige basis worden samengesteld op basis van administratieve gegevens die door de autoriteiten van de lidstaten worden bewaard, gedocumenteerd en opgeslagen in een archief met gecontroleerde toegang. Erkende contractanten kunnen met toestemming gegevens uit dergelijke archieven halen. Wanneer er geen specifieke toestemming kan worden verkregen, kunnen gegevens volledig worden geanonimiseerd met versleutelde persoonlijke identificatiemiddelen om te voldoen aan de AVG-verordening. In verschillende landen zijn dergelijke gegevensopslagplaatsen opgezet. Wanneer toegang echter een probleem blijft als gevolg van uiteenlopende interpretaties van de privacyregels, zou een nationaal initiatief op overheidsniveau overeenkomsten en systemen moeten bevorderen die in relatief korte tijd gegevens voor onderzoeksdoeleinden kunnen leveren.

*Analytische datasets creëren*

Wanneer er bezorgdheid blijft bestaan over de vertrouwelijkheid van persoonsgegevens, zou moeten worden overwogen om gegevenslaboratoria op te richten. Beoordelaren die met administratieve datasets werken, zouden alleen toegang tot gegevens moeten krijgen op veilige locaties, waar de toegang tot gegevens streng wordt gemonitord en gecontroleerd. De gegevens zouden ter plaatse moeten worden verwerkt en geanalyseerd, en alleen de resultaten van analyses zouden de locatie mogen verlaten.

*Gegevenslaboratoria opzetten*

### 4.4. Op weg naar meer toekomstgerichte benaderingen

Een gemeenschappelijk kenmerk van het tot op heden beperkte aantal CIE's dat is uitgevoerd betreffende uit het ESF-gefinancierde interventies, is dat deze eerder retrospectief dan prospectief van aard waren. Dit houdt in dat

deskundige beoordelaren opdracht kregen evaluaties uit te voeren van interventies die waren ontwikkeld zonder rekening te houden met evaluaties, en waar in sommige gevallen weinig of geen planning voor een effectonderzoek had plaatsgevonden. Daarom moesten beoordelaren op tijdrovende, dure en suboptimale manieren gegevensbronnen opstellen, aan de hand van gegevens die toevallig beschikbaar waren, in plaats van op basis van gegevensbronnen die waren samengesteld met het oog op effectonderzoek.

Bij een prospectieve benadering zouden beoordelaren daarentegen zo snel mogelijk een CIE moeten plannen en zou het mogelijk zijn om (nieuwe of bestaande) interventies te beïnvloeden, op vaak zeer subtiele manieren, waardoor zij vatbaarder voor een CIE zouden zijn. De planning vooraf van een CIE kan het verschil maken tussen de mogelijkheid om een grondige evaluatie of juist helemaal geen evaluatie te kunnen uitvoeren. Het inschakelen van goed opgeleid intern personeel of het in dienst nemen van externe deskundige contractanten in een vroeg stadium van een interventie of wanneer financieringsbesluiten worden genomen, houdt in dat:

- het naar behoren bijhouden van een administratie kan worden opgenomen in de uitvoering van programma's en interventies;
- de vereiste gegevensbronnen in een vroeg stadium kunnen worden geïdentificeerd en problemen op het gebied van toegang tot gegevens en gegevensbescherming tijdig worden aangepakt;
- de verzameling van referentiegegevens in detail kan worden gepland en zo nodig enquêtes kunnen worden georganiseerd;
- praktische problemen in verband met de wijze waarop deelnemers voor interventies worden geworven, op zodanige wijze kunnen worden aangepakt dat de wervingsprocedures beter aansluiten op een rigoureuze evaluatie.

De betrokkenheid van beoordelaren die zijn opgeleid in CIE-methoden (of het nu gaat om interne beoordelaren van de beheerautoriteit of externe deskundigen) bij de ontwikkeling van nieuwe ESF+-interventies of bij besluiten over bestaande interventies, zou het mogelijk maken om aan het begin van de programmeringsperiode met de planning van effectonderzoeken te beginnen en aanzienlijke voordelen te behalen voor empirisch onderbouwde beleidsbeslissingen.

#### 4.5. Het toepassingsgebied van CIE's uitbreiden

Een laatste, zeer belangrijk punt is de mate waarin CIE's het beleid bestrijken, en dit punt omvat de volgende aspecten: het behandelde beleid, de geanalyseerde uitkomsten en de volledigheid van de interpretatie van de bevindingen.

In de periode 2014-2020 was de verspreiding van CIE's vrijwel uitsluitend gericht op actief arbeidsmarktbeleid en effecten in verband met de arbeidsstatus. Dit kan waarschijnlijk worden verklaard door het feit dat administratieve gegevens over werkgelegenheid over het algemeen gemakkelijker toegankelijk zijn en dat deze gegevens de belangrijkste uitkomstvariabele vormen voor de analyse van veel beleidsmaatregelen op het gebied van arbeidsmarktactivering.

*Prospectieve methoden voorzien de betrokkenheid van beoordelaren vanaf het begin*

*Meer aandacht voor CIE's in het onderwijsbeleid en het beleid inzake sociale inclusie*

Het onderwijsbeleid en het beleid op het gebied van sociale inclusie zijn, ondanks het belang ervan in de ESF-strategie en de toewijzing van middelen, niet of slechts zeer zelden onderworpen aan een CIE; enkele van de weinige beschikbare voorbeelden uit Spanje, Polen en Portugal worden in tekstkader 20 besproken.

Om dit tekort aan CIE's weg te werken en de kennis van deze belangrijke beleidsterreinen te verbeteren, is een reeks gecombineerde initiatieven nodig:

- het vooraf plannen en voorbereiden van CIE's op het gebied van sociale inclusie- of onderwijsmaatregelen, zodat de beschikbaarheid van gegevens tijdig kan worden beoordeeld; de nodige samenwerking tussen verschillende actoren op gang brengen om die beschikbaarheid te vergroten; en de behandelings- en controlegroep in een vroeg stadium identificeren.
- het bevorderen van de betrokkenheid van de gegevenseigenaren bij de CIE-opzet; een bredere betrokkenheid bij CIE's van de instanties die verantwoordelijk zijn voor onderwijs en sociaal beleid, wanneer deze van de beheerautoriteit verschillen, zou een sterkere en meer algemene inzet op gang kunnen brengen voor evaluaties en het beschikbaar stellen van gegevens;
- institutionele overeenkomsten en software-instrumenten stimuleren om administratieve gegevens over sociale omstandigheden en onderwijs bruikbaar te maken voor CIE's; Daartoe kan de betrokkenheid van het nationale bureau voor de statistiek als "brug" en facilitator van belang zijn, aangezien deze bureaus over de vereiste vaardigheden beschikken en al veel administratieve datasets voor de productie van nationale statistieken uitwerken;
- het ondersteunen van de verzameling van relevante gegevens van deelnemers in de behandelings- en de controlegroep aan het begin van de interventie om te kunnen beschikken over een consistente gegevensreeks "voor en na" de interventie. Wanneer administratieve gegevens niet beschikbaar of relevant zijn, kan een specifieke enquête de nodige gegevens opleveren, maar deze moet aan het begin van de interventie van start gaan en ook tot de controlegroep gericht zijn om de nodige vergelijkingen te kunnen maken.

*Andere actoren  
betrekken en  
geschikte  
gegevensbronnen  
ontwikkelen*

### **Tekstkader 20 Voorbeelden van evaluaties op onderwijsgebied**

#### **In Asturië uitgevoerde ESF-maatregelen tegen voortijdig schoolverlaten<sup>98</sup>**

In het kader van een algemene evaluatie van het operationeel programma "Principado de Asturias ESF 2014-2020" is een specifieke effectbeoordeling van het *Diversificación Curricular y de Mejora del Aprendizaje y del Rendimiento* (PMAR) uitgevoerd. PMAR is een ESF-maatregel die in het kader van IP10.i wordt uitgevoerd om voortijdig schoolverlaten tegen te gaan. In het kader van PMAR werden leerlingen verdeeld in specifieke groepen (van 8 tot 15 leerlingen, maar met een andere omvang als er sprake is van bijzondere omstandigheden), om lessen te volgen op het gebied van taalkunde, sociale wetenschappen, wetenschap en wiskunde, evenals vreemde talen (de overige vakken worden in de hoofdklas onderwezen). De maatregel werd uitgevoerd in het studiejaar 2016-2017, in het tweede en derde jaar van de middelbare school (*Educación Secundaria Obligatoria*, ESO). Leerlingen die op een bepaald moment ten minste één keer waren blijven zitten en leerlingen die het eerste jaar ESO hadden afgerond, maar niet geschikt waren

<sup>98</sup> Zie [Diaz J.M. et al., 2019](#)

om over te gaan, kwamen in aanmerking voor deelname. De uiteindelijke selectie vindt plaats op basis van de individuele beoordeling van de leerlingen (academisch en psychopedagogisch) door de onderwijsteams.

In totaal namen 1 053 leerlingen deel, waarvan 512 uit het tweede jaar van de ESO en 541 uit het derde jaar. Deze leerlingen werden vergeleken met andere leerlingen met leerproblemen<sup>99</sup>, om controlepersonen te selecteren die wat betreft de criteria voor het al dan niet in aanmerking komen, vergelijkbaar waren met de te behandelen leerlingen (in totaal 3 852 personen).

Er werd een matchingmethode toegepast met gebruik van de administratieve gegevens uit de SAUCE-databank (*Sistema Informático para la Administración Unificada de Centros Educativos*), die in geanonimiseerde vorm door het ministerie van Onderwijs werden verstrekt. Deze gegevens werden gebruikt om de propensity-score<sup>100</sup> te berekenen en de uitkomstvariabele, namelijk de succesvolle overgang naar het volgende studiejaar, te meten. Uit de evaluatie is gebleken dat de interventies een positief en statistisch significant effect hadden: in de behandelingsgroep hadden de leerlingen uit het tweede jaar een slaagpercentage dat 18 procentpunten hoger lag dan dat van de controlegroep, terwijl voor de behandelde leerlingen uit het derde jaar het verschil ongeveer 16 procentpunten bedroeg. In beide groepen was het effect positiever voor meisjes dan voor jongens.

### **ESF-maatregelen ter verbetering van het beroepsonderwijs in Podlaskie<sup>101</sup>**

In de studie werden de effecten onderzocht van het project “Goede baan — geweldig leven”, dat gericht was op de bevordering van beroepsonderwijs en -opleiding in de provincie Podlaskie in Polen. Het project was gericht op medewerkers van lagere secundaire scholen, leerlingen en ouders, en er werd een reeks activiteiten uitgevoerd voor zowel de instellingen als de mensen (op maat gesneden marketing- en communicatiestrategieën voor scholen, opleiding voor leerkrachten, samenwerking met bedrijven en lokale bedrijfsleiders, mentorschap voor leerlingen en ouders, beroeps- en onderwijsadvies voor leerlingen). Het project werd uitgevoerd tussen 2017 en 2019.

De behandelde populatie bestond uit leerlingen die deelnamen aan het geëvalueerde project. Van de 9 403 leerlingen werden er 1 500 geselecteerd door middel van gestratificeerde willekeurige steekproeven, en namen 200 leerlingen deel aan de enquête van de studie. De informatie over leerlingen die aan het project deelnamen, werd verkregen uit monitoringgegevens. De controlegroep moest leerlingen omvatten met kenmerken die vergelijkbaar waren met die van de behandelde groep, maar die niet aan het project deelnamen. Er werd een controlegroep gepland die ongeveer dubbel zo groot was als de behandelingsgroep. De leden van de controlegroep werden willekeurig geselecteerd uit het nationale bevolkingsregister (PESEL — een register dat aan elke Poolse burger een individueel nummer toekent). Door problemen als gevolg van covid was de geplande methode echter niet uitvoerbaar. Uiteindelijk werd de sneeuwbaltechniek (kettingverwijzing) gebruikt als primaire steekproefmethode, wat leidde tot 401 afgenomen interviews, waarvan er 384 in de studie werden gebruikt.

Gegevens over de uitkomstvariabelen<sup>102</sup> voor zowel de behandelings- als de controlegroep werden verzameld aan de hand van enquêtes op basis van vragenlijsten, met behulp van persoonlijke, telefonische of webinterviews. In de studie werd de propensity-score gebruikt als contrafeitelijke methode voor de beoordeling van het effect (alleen op algemeen niveau, niet voor specifieke subgroepen). De in de studie uitgevoerde contrafeitelijke onderzoeken leverden geen sluitende conclusies op. De statistische significantie van de resultaten was onbevredigend.

<sup>99</sup> Deze werden gevonden onder leerlingen die in het studiejaar 2015-2016 in hun eerste en tweede jaar secundair onderwijs zaten, die het jaar met een of meer onvoldoendes afronden en die in hun schoolcarrière ten minste één keer waren blijven zitten.

<sup>100</sup> De variabelen op grond waarvan de twee groepen moesten worden gematcht, waren: schoolprestaties tijdens het voorafgaande studiejaar, soort school (openbaar of particulier), inschrijving voor een andere aanvullend vak tijdens het voorafgaande studiejaar (ja of nee), geslacht, gemiddeld inkomen in de woongemeente, woonplaats, geboorteland (Spanje of de rest van de wereld).

<sup>101</sup> Zie [Zub M. et al., 2020](#)

<sup>102</sup> De uitkomstvariabelen (op basis van de antwoorden in de enquête) omvatten: inschrijving bij een technische school op het moment van het onderzoek, inschrijving bij een eerstegraadsvakschool op het moment van het onderzoek, inschrijving bij hetzij een technische school, hetzij een eerstegraadsvakschool op het moment van het onderzoek, aanmelding bij een middelbare school, aanmelding bij een technische school, aanmelding bij een eerstegraadsvakschool, aanmelding bij een school voor beroepsopleiding (technische school of eerstegraadsvakschool), waarbij de school in zeer hoge mate aan de verwachtingen van de leerling voldoet en de studierichting absoluut aansluit bij de interesses van de leerling.



## Evaluatie van het stelsel van studiebeurzen voor het hoger onderwijs voor kansarme studenten in Portugal<sup>103</sup>

In een tweede, nog niet officieel gepubliceerde evaluatie wordt een programma voor studiebeurzen voor studenten uit huishoudens met een laag inkomen onderzocht. Het programma heeft tot doel hun toegang tot het hoger onderwijs te bevorderen en het aantal inschrijvingen van studenten uit deze groep te verhogen. De studiebeurzen worden ondersteund door het ESF (sinds de periode 2007-2013) in de regio's Noord- en Midden-Portugal en Alentejo, en door de staat in de overige regio's. In aanmerking komen studenten met middelen onder een specifieke drempel (tussen 6 800 en 7 900 EUR inkomen per hoofd van de bevolking) en een minimumaantal studiepunten dat in het voorgaande studiejaar is behaald voor studenten die niet in hun eerste academische jaar zitten. De voorschriften om in aanmerking te komen voor een studiebeurs zijn hetzelfde voor inschrijving bij openbare als bij particuliere universiteiten. Het bedrag van de beurs is evenredig met het inkomen van het huishouden.

Sinds 2011 telt het programma jaarlijks ongeveer 70 000 studenten, maar de evaluatie bestrijkt de periode vanaf 2012, omdat vanaf dat jaar gegevens beschikbaar waren. Bovendien richt de evaluatie zich op studenten die voor het eerst een beurs aanvragen en ingeschreven zijn in het eerste jaar van een universitaire opleiding (bachelor of master); het recht op een beurs wordt alleen bepaald aan de hand van het inkomenscriterium<sup>104</sup>.

Er werden twee administratieve datasets gebruikt: gegevens over studenten die zich tussen 2012 en 2018 aanmeldden (verstrekkt door het directoraat-generaal voor het hoger onderwijs in Portugal (DGES)), die via een unieke identificatiecode werden samengevoegd met een andere dataset met informatie over academische carrières en doorstroming (verstrekkt door het directoraat-generaal voor statistiek en wetenschappen (DGEEC)). De uiteindelijke steekproef die wordt geanalyseerd, bestaat uit 156 002 studenten, waarvan er 130 602 werden behandeld en 25 400 niet. Op grond van de inkomensdrempel (de lopende variabele) was de beoordelaar in staat een regressie-discontinuïteitsanalyse toe te passen en werden verschillende uitkomstvariabelen gemeten: op korte termijn, het percentage studenten dat aan het eind van het eerste jaar nog ingeschreven was, en de aan het eind van het eerste jaar behaalde studiepunten; op langere termijn, de waarschijnlijkheid van het behalen van het diploma, het behaalde eindcijfer en het aantal jaren waarin het diploma wordt behaald.

In de beginperiode vertoont het programma voor studiebeurzen positieve effecten op het gebied van het aantal inschrijvingen, maar niet wat studiepunten betreft; op de langere termijn studeren er meer behandelde studenten af, in minder tijd en met hogere cijfers dan de controlegroep. De effecten waren groter bij mannelijke studenten, studenten die in gebieden wonen die onder het ESF vallen en studenten die zijn ingeschreven bij masterstudies en staatsuniversiteiten.

Er is al eerder gewezen op het nut en de verklarende capaciteit van de meta-evaluaties. Een algemene toename van CIE's en bijgevolg een mogelijke toename van meta-evaluaties zou het debat over specifieke maatregelen versterken, waarbij tegelijkertijd verschillende beheerautoriteiten en lidstaten zouden worden betrokken. Zoals hierboven vermeld, moet de methodologische complexiteit van deze evaluaties worden overwonnen door een betere planning van CIE's op nationaal en internationaal niveau.

*Het aantal meta-evaluaties verhogen*

Bij ESF-interventies meten CIE's over het algemeen effecten met kwantitatieve of binaire variabelen (ja/nee), zoals loon of arbeidsstatus. Soft outcomes, die verband houden met het zelfbeeld en de capaciteit van de deelnemers, zijn echter belangrijke graadmeters voor het succes van de interventies. Zo zijn "inzetbaarheid", zelfvertrouwen of verworven vaardigheden voorwaarden om een baan te vinden of actief te zijn op de arbeidsmarkt. In veel contexten zouden de analyses van soft outcomes zelfs

*Soft outcomes in aanmerking nemen*

<sup>103</sup> Zie Guthmuller S., Meroni E.C., nog niet gepubliceerd.

<sup>104</sup> Er wordt een analyse van de volledige steekproef van studenten (niet alleen die van het eerste jaar) verwacht, maar deze is nog niet voltooid.

nuttiger zijn dan cijfers van de uitstroom uit het arbeidsproces en zouden zij een licht werpen op het matchen van interventies en individuele behoeften.

CIE's kunnen op deze soft outcomes worden toegepast op voorwaarde dat de informatie vóór en na de interventie in zowel de behandelings- als de controlegroep wordt verzameld overeenkomstig de interventielogica. Deze onderzoeken moeten in een vroeg stadium worden gepland en alle relevante actoren, met name de begunstigen die rechtstreeks en voortdurend in contact staan met de deelnemers, moeten hierbij worden betrokken. Zie tekstkader 21 voor een voorbeeld van een CIE waarin de effecten op soft outcomes worden beoordeeld.

### **Tekstkader 21 Voorbeeld van de beoordeling van effecten op soft outcomes in Duitsland**

#### **Kunnen regelingen voor het scheppen van banen de sociale integratie en het welzijn van langdurig werklozen verbeteren<sup>105</sup>?**

In de studie worden de effecten onderzocht van een regeling voor het scheppen van banen die gericht is op een kwetsbare groep personen, namelijk langdurig werklozen die al ten minste vier jaar een uitkering ontvangen en die een gezondheidsbeperking of kinderen, of beide, hebben. Het programma liep van 2015 tot 2018 en omvatte ongeveer 20 000 deelnemers. In het kader van de maatregel werden reguliere arbeidsovereenkomsten tot 36 maanden lang voor 30 werkuren per week gesubsidieerd (voornamelijk met openbare werkgevers of liefdadigheidsorganisaties).

Aangezien de maatregel van gesubsidieerd werk, die maximaal 36 maanden duurt, uitdrukkelijk was bedoeld om de sociale integratie van de doelgroepen te bevorderen, probeerden de auteurs de effecten op dat gebied te beoordelen.

Voor de analyse werd een rijke dataset gebruikt waarin administratieve gegevens (de Duitse geïntegreerde beroepsbiografieën) werden aangevuld met een panelonderzoek onder de deelnemers aan het programma en de personen uit de controlegroep. Dankzij het panelonderzoek was de verzameling van informatie over subjectieve maatstaven van levenskwaliteit mogelijk: persoonlijke tevredenheid, geestelijke gezondheid, maatschappelijke verbondenheid en sociale status. Om informatie te verzamelen over deze maatstaven, werd gebruik gemaakt van Likert-schalen, die ook in onderzoeken (zoals de nationale enquête PASS) werden gebruikt. Op die manier konden vergelijkbare bevindingen worden verkregen. Dit waren de belangrijkste stappen van het onderzoek: de te behandelen en niet te behandelen personen werden geïdentificeerd en gematcht op grond van de administratieve gegevens; de daaruit voortvloeiende behandelingsgroep en controlegroep werden geïnterviewd (in drie rondes); de definitieve dataset werd opgesteld, de gegevens werden opgeschoond en gevallen met ontbrekende gegevens werden verwijderd. De steekproef bestond uit 2 531 gematchte paren in fase 1; 1 191 in fase 2, en 450 in fase 3 (op een totale steekproef van circa 62 000 mensen, waarvan 12 400 in de behandelingsgroep en 49 600 in de controlegroep).

Op basis van propensity score matching en het meten van de effecten na 7, 18 en 29 maanden vanaf de toetreding tot het programma, stelden de auteurs vast dat de interventies positieve effecten hadden op de maatstaven van welzijn, maar in verschillende mate: zo was de persoonlijke tevredenheid aanzienlijk toegenomen, terwijl de sociale status slechts matig was verbeterd. De effecten nemen echter in de loop van de programma's af en dit kan worden verklaard door een toename van zowel het aantal deelnemers dat het programma verlaat als het aantal personen uit de controlegroep dat een baan vindt. Voor de meest kwetsbare mensen waren de positieve effecten groter.

<sup>105</sup> Zie [Ivanov B. et al., 2020](#)

Zoals aan het begin van deze leidraad werd benadrukt, is CIE een krachtig analytisch instrument om te beoordelen “wat” de ESF+-interventie heeft voortgebracht. Om te begrijpen “hoe” of “waarom” de gemeten effecten zich hebben voorgedaan, moeten andere instrumenten worden gebruikt.

*Integratie van  
CIE's en  
theoriegebaseerde  
evaluaties*

In sommige gevallen, wanneer de interventies welbekend zijn en al grondig zijn geëvalueerd, volstaan een relatief beperkte analyse van de uitvoering en enkele interviews met de begunstigden om de belangrijkste elementen te identificeren die aan de basis liggen van de gemeten effecten. Bij complexere, minder bekende of innovatieve interventies moet het CIE vaak worden aangevuld met andere evaluaties.

Uit een vóór het CIE uitgevoerde evaluatie van de uitvoering of het proces kan blijken hoe de interventie is verlopen, welke problemen zich bij de uitvoering hebben voorgedaan en hoe de verschillende actoren hebben samengewerkt om de interventie te doen slagen. De bevindingen van een dergelijke evaluatie kunnen als input dienen voor het CIE, en aangeven welke effecten moeten worden onderzocht, hoe de controlegroep moet worden samengesteld met betrekking tot de deelnamecriteria, en wat de juiste timing voor een CIE is.

Een theoriegebaseerde evaluatie (TBE) kan tegelijkertijd met het CIE worden uitgevoerd. De twee evaluaties zouden elkaar kunnen aanvullen: de theoriegebaseerde evaluatie geeft uitleg over de contextuele, sociale en individuele mechanismen die de effecten hebben beïnvloed en mogelijk hebben gemaakt, terwijl het CIE netto-effecten berekent en verduidelijkt in welke mate de interventie significante uitkomsten heeft opgeleverd. In een effectonderzoek wordt de integratie van CIE en TBE aanbevolen om een uitgebreid overzicht te geven van de effecten en de oorzaken ervan.

# Glossaria

## Afkortingen

AAMB	Actief arbeidsmarktbeleid
KBA	Kosten-batenanalyse
CIE	Contrafeitelijk effectonderzoek (Counterfactual impact evaluation)
GB-verordening	Verordening gemeenschappelijke bepalingen
DG EMPL	Directoraat-generaal Werkgelegenheid, Sociale Zaken en Inclusie
DG REGIO	Directoraat-generaal Regionaal beleid
DID	Difference-in-difference/s
EC	Europese Commissie
EP	Evaluatieplan
EFRO	Europees Fonds voor regionale ontwikkeling
ESF	Europees Sociaal Fonds
EU	Europese Unie
AVG	Algemene verordening gegevensbescherming (Verordening (EU) 2016/679)
IB	Intermediaire instantie
IEB	Geïntegreerde beroepsbiografieën (Integrierten Erwerbsbiografien)
IV	Instrumentele variabele
JLD	Jobseeker longitudinal dataset (Ierse longitudinale dataset van werkzoekenden)
AKE	Arbeidskrachtenenquête
BA	Beheerautoriteit
LS	Lidstaat
Ngo	Niet-gouvernementele organisatie
OP	Operationeel programma
ODA	Openbare dienst voor arbeidsvoorziening
PSM	Propensity score matching
RCT	Gerandomiseerd onderzoek met een controlegroep (Randomised controlled trial)
RDD	Regressie-discontinuïteitsanalyse (Regression discontinuity design)
Kmo's	Kleine en middelgrote ondernemingen
YEI	Jongerenwerkgelegenheidsinitiatief (Youth Employment Initiative)

## Definities

Begrip	Definitie
Referentiegegevens	Gegevens over variabelen die zijn gemeten voordat een eenheid (persoon of onderneming) aan een interventie deelneemt. In veel gevallen zullen aan de behandeling voorafgaande metingen van de interventieresultaten worden verzameld voor zowel de behandelings- als de controlegroep.
Begunstigde	Volgens artikel 2 van Verordening (EU) 2021/1060 (GB-verordening) is een begunstigde “een publiek- of privaatrechtelijke instantie, een entiteit met of zonder rechtspersoonlijkheid of een natuurlijke persoon die belast is met het opzetten, of met het opzetten en uitvoeren, van concrete acties”. Waar het gaat om financieringsinstrumenten (soms gebruikt voor microkredieten voor zelfstandige arbeid in het kader van het ESF+) is een begunstigde de instantie die het holdingfonds uitvoert of, bij ontstentenis van een holdingfondsstructuur, de instantie die het specifieke fonds uitvoert, of de beheerautoriteit, indien het financieringsinstrument door deze wordt beheerd.
Controlegroep	Een groep personen, ondernemingen of andere eenheden die zo vergelijkbaar mogelijk is met de behandelingsgroep, maar die niet behandeld wordt en waaruit contrafeitelijke schattingen van de resultaten worden verkregen. Strikt genomen verwijst de term “controlegroep” naar een experimentele context (zie “randomisatie” hieronder) en verwijst de term “vergelijkingsgroep” naar een quasi-experimentele context, maar in de praktijk worden beide termen door elkaar gebruikt.
Contrafeitelijke analyse	Een vergelijking tussen wat er daadwerkelijk is gebeurd en wat er zonder de interventie zou zijn gebeurd, met het oog op de resultaten. Aangezien het verschil tussen de feitelijke en de contrafeitelijke resultaten het causale effect van de interventie bepaalt, omvat de contrafeitelijke analyse alle benaderingen die erop gericht zijn het aandeel van de waargenomen verandering dat aan de interventie in kwestie kan worden toegeschreven, te beoordelen.
Difference-in-differences (DID)	In zijn eenvoudigste vorm betekent dit dat het verschil in een resultaat vóór en na behandeling in een controlegroep wordt afgetrokken van hetzelfde verschil dat is waargenomen in een behandelde groep om een schatting van het effect van een interventie te verkrijgen. Effecten die op basis van de DID-methode worden berekend, worden doorgaans binnen een regressiekader afgeleid.
Doeltreffendheid	Effectiviteit verwijst naar “verwezenlijking van de doelstellingen” en wordt beoordeeld aan de hand van een vergelijking tussen wat is bereikt en wat was gepland (of een referentiesituatie) of een vergelijking tussen wat na de actie is waargenomen en wat zonder de actie zou zijn gebeurd (contrafeitelijke situatie).
Efficiëntie	Efficiëntie wordt gedefinieerd als het verkrijgen van een bepaalde output tegen de minimale kostprijs of, op gelijkwaardige wijze, als het maximaliseren van de output voor een bepaald niveau van middelen. Dit kan worden vastgesteld aan de hand van een kosten-batenanalyse of een kosten-effectiviteitsanalyse.
Evaluatieplan	Overeenkomstig artikel 44 van Verordening 2021/1060 (GB-verordening), moet de lidstaat of de beheerautoriteit een evaluatieplan opstellen dat betrekking kan hebben op meer dan één programma en dat uiterlijk een jaar na het besluit tot goedkeuring van het programma bij het monitoringcomité moet worden ingediend.
Externe evaluatie	Evaluatie die wordt uitgevoerd door een externe en onafhankelijke beoordelaar op basis van een aanbestedingsprocedure.
Effecten	In het kader van een CIE verwijzen effecten naar netto-effecten, gedefinieerd als het verschil tussen de gemiddelde resultaten van de behandeling en de contrafeitelijke resultaten. Voor de toepassing van deze leidraad, worden de termen “effecten” en “netto-effecten” door elkaar gebruikt.
Contrafeitelijk effectonderzoek	Een type effectonderzoek waarbij wordt getracht de causale effecten van interventies vast te stellen door de gemiddelde contrafeitelijke resultaten te schatten en deze af te trekken van de gemiddelde waargenomen resultaten bij behandelde eenheden. Schattingen van contrafeitelijke resultaten worden doorgaans verkregen van een controlegroep die zorgvuldig is geselecteerd om zo vergelijkbaar mogelijk te zijn met de behandelde groep.

Begrip	Definitie
Instrumentele-variabelenmethode (IV)	Bij de toepassing van deze methode moet de selectie voor deelname aan de behandeling ten minste gedeeltelijk worden bepaald door een exogene factor (of een exogeen instrument) die geen verband houdt met andere resultaten dan die via de behandeling worden verkregen. De exogene factor is dus van invloed op de deelname, maar niet rechtstreeks op de resultaten.
Interne evaluatie	Intern uitgevoerde evaluatie, d.w.z. rechtstreeks in opdracht van een onafhankelijke overheidsinstantie of -afdeling (van de beheerautoriteit of de intermediaire instantie) zonder aanbestedingsprocedure of in de vorm van een uitgebreid monitoring- en analyseproces.
Interventies	Deze verwijzen in het algemeen naar concrete acties in operationele programma's van het ESF of naar uit het ESF medegefinancierde projecten.
Matching	Dit is een methode waarbij interventie- en controlesteekproeven op basis van hun waargenomen kenmerken worden gematcht.
Niet-gerandomiseerde of quasi-experimentele opzet	Benaderingen voor contrafeitelijk effectonderzoek waarbij controlegroepen worden samengesteld met behulp van andere methoden dan randomisatie.
Uitkomst	De waarschijnlijke of bereikte effecten op korte en middellange termijn van de outputs van een interventie <sup>106</sup> . Een uitkomst betekent ongeveer hetzelfde als "resultaat", maar wordt vaker gebruikt in effectonderzoeken.
Output	Output wordt beschouwd als alles wat rechtstreeks wordt geproduceerd/geleverd door de uitvoering van een concrete actie van het ESF, gemeten in fysieke of monetaire eenheden. Outputs worden voornamelijk gemeten aan de hand van het aantal ondersteunde personen, ondersteunde entiteiten, geleverde goederen en diensten en uitgevoerde projecten.
Deelnemers	Een deelnemer is een natuurlijke persoon die rechtstreeks voordeel haalt uit een concrete actie, zonder verantwoordelijk te zijn voor het opstarten of uitvoeren van die actie. In de context van het ESF verwijst de term naar personen die door ESF-interventies worden ondersteund.
Procesevaluatie	De procesevaluatie is gericht op de uitvoering van het programma, met inbegrip van, maar niet beperkt tot, de wijze waarop diensten worden verleend, de verschillen tussen de geplande doelgroep en de groep aan wie de dienst daadwerkelijk is verleend, de toegang tot het programma en de beheerpraktijken.
Propensity score matching (PSM)	PSM houdt in dat voor de gehele steekproef (behandelingseenheden en potentiële controle-eenheden) een statistisch model wordt geraamd dat een geschatte waarschijnlijkheid van deelname voor elke persoon of onderneming oplevert, ongeacht of deze persoon daadwerkelijk deelneemt of niet. Op basis van de propensity-score worden behandelde personen of ondernemingen vervolgens met één niet-behandelde persoon of onderneming, dan wel met meerdere niet-behandelde personen of ondernemingen gematcht.
Randomisatie	Dit is een methode waarbij leden van een doelgroep willekeurig worden toegewezen aan een reeks behandelingen of aan controleomstandigheden. Randomisatie zorgt ervoor dat de groepen statistisch in alle aspecten gelijkwaardig zijn op het moment dat zij worden gerandomiseerd.
Regressie-discontinuïteitsanalyse (RDD)	Deze methode kan worden toegepast wanneer de toegang tot een interventie wordt bepaald door een grenswaarde in een continue beoordeling, schaal of meting. Bij deze benadering wordt gebruik gemaakt van het feit dat de eenheden die zich in de onmiddellijke nabijheid van de grenswaarde bevinden, onderling zeer vergelijkbaar zijn, met als enig verschil dat de eenheden aan de ene kant van de grenswaarde wel deelnemen, en die aan de andere kant niet. De resultaten voor eenheden boven en onder de grenswaarde kunnen met elkaar worden vergeleken om de effecten van de interventie te bepalen.
Relevantie	Met relevantie wordt verwezen naar de geschiktheid van de expliciete doelstellingen van een interventie met betrekking tot de sociaal-economische problemen die met de

<sup>106</sup> Zie [OESO, 2010](#)

<b>Begrip</b>	<b>Definitie</b>
	<p>interventie moeten worden opgelost<sup>107</sup>.</p> <p>De effecten van een concrete actie op de deelnemers of entiteiten, bijvoorbeeld wat betreft hun arbeidsparticipatie, inkomen, scores in gestandaardiseerde onderwijstests, winst enz. De effecten kunnen op korte of langere termijn worden gemeten. In de context van het ESF+ wordt bij het meten van resultaatindicatoren de periode onmiddellijk na deelname (vier weken) gezien als de korte termijn, terwijl de langere termijn verwijst naar zes maanden daarna<sup>108</sup>. In het kader van het effectonderzoek wordt met langere termijn echter vaak een langere periode bedoeld, namelijk 24 maanden of meer na de interventies.</p>
Resultaat	
Behandelingsgroep (ook behandelde groep)	<p>Een groep personen, ondernemingen of andere eenheden die baat hebben bij of deelnemen aan een interventie (dit kan het aanbod van de behandeling of de daadwerkelijke ontvangst ervan zijn).</p>

---

<sup>107</sup> Zie [Europese Commissie, 2013](#)

<sup>108</sup> Zie [Europese Commissie, 2021a](#) en [Europese Commissie, 2018](#)

## Bibliografie

- Ashenfelter, O. (1978), "Estimating the effect of training programmes on earnings", *Review of Economics and Statistics*, 6, blz. 47-57, <https://www.jstor.org/stable/1924332?origin=crossref>
- Axdorph E, Egebark J., Lundström T., Özcan G. (2019), *Effekter av förstärkta förmedlingsinsatser för unga arbetssökande — Resultat från utvärderingen av Ung framtid*, Arbetsförmedlingen.
- Baran J. et al. (2016), *Badanie efektów wsparcia zrealizowanego na rzecz osób młodych w ramach Programu Operacyjnego Wiedza Edukacja Rozwój. I Raport Wskaźnikowy*, <http://files.evaluationhelpdesk.eu/Evaluations/PLE49.pdf>
- Baran J. et al. (2017), *Badanie efektów wsparcia zrealizowanego na rzecz osób młodych w ramach Programu Operacyjnego Wiedza Edukacja Rozwój, II Raport wskaźnikowy*, <http://files.evaluationhelpdesk.eu/Evaluations/PLE196.pdf>
- Baran J. et al. (2018), *Badanie efektów wsparcia zrealizowanego na rzecz osób młodych w ramach Programu Operacyjnego Wiedza Edukacja Rozwój, III Raport wskaźnikowy*, <http://files.evaluationhelpdesk.eu/Evaluations/PLE197.pdf>
- Baran J. et al. (2018a), *Badanie efektów wsparcia zrealizowanego na rzecz osób młodych w ramach Programu Operacyjnego Wiedza Edukacja Rozwój, II Raport Tematyczny*, <http://files.evaluationhelpdesk.eu/Evaluations/PLE221.pdf>
- Bazzoli M., De Poli S., Rettore E., Schizzerotto A. (2018), "Are Vocational Training Programmes Worth Their Cost? Evidence from a Cost-Benefit Analysis", *Politica Economica/Journal of Economic Policy*, Vol. XXXIV(3), blz. 215-240, <https://www.rivisteweb.it/doi/10.1429/92119>
- Bloom, H. S. (1995), "Minimum detectable effects: A simple way to report the statistical power of experimental designs", *Evaluation Review*, 8(2), blz. 225-246, <https://cpb-us-e2.wpmucdn.com/sites.uci.edu/dist/1/1159/files/2021/03/Bloom-MDES-Eval-Rev-1995-Bloom.pdf>
- Bloom, H. S. (2009), "Modern regression discontinuity analysis", *MDRC Working Papers on Research Methodology*, New York: MDRC, <https://www.mdrc.org/publication/modern-regression-discontinuity-analysis>
- Boockmann B. et al. (2017), *Evaluation des ESF-Bundesprogramms zur Eingliederung langzeitarbeitsloser Leistungsberechtigter nach dem SGB II auf dem allgemeinen Arbeitsmarkt. Im Auftrag des Bundesministeriums für Arbeit und Soziales*, <http://files.evaluationhelpdesk.eu/Evaluations/DEE189.pdf>
- Boockmann B. et al. (2018), *Evaluation des ESF-Bundesprogramms zur Eingliederung langzeitarbeitsloser Leistungsberechtigter nach dem SGB II auf dem allgemeinen Arbeitsmarkt. Im Auftrag des Bundesministeriums für Arbeit und Soziales*, <http://files.evaluationhelpdesk.eu/Evaluations/DEE52.pdf>
- Boockmann B. et al. (2019), *Evaluation des ESF-Bundesprogramms zur Eingliederung langzeitarbeitsloser Leistungsberechtigter nach dem SGB II auf dem allgemeinen Arbeitsmarkt. Im Auftrag des Bundesministeriums für Arbeit und Soziales*, <http://files.evaluationhelpdesk.eu/Evaluations/DEE94.pdf>
- Boockmann B. et al. (2021), *Evaluation des ESF-Bundesprogramms zur Eingliederung langzeitarbeitsloser Leistungsberechtigter nach dem SGB II auf dem allgemeinen Arbeitsmarkt. Im Auftrag des Bundesministeriums für Arbeit und Soziales*, <http://files.evaluationhelpdesk.eu/Evaluations/DEE202.pdf>
- Borik V. et al. (2015), *The net effects of graduate work experience and the promotion of self-employment. Technical Report*, <http://files.evaluationhelpdesk.eu/Evaluations/SKE8.pdf>



- Bratti M. et al. (2018), *Vocational Training for Unemployed Youth in Latvia: Evidence from a Regression Discontinuity Design*, IZA DP No. 11870, <http://ftp.iza.org/dp11870.pdf>
- Bratu C. et al. (2014), *Knowledge gaps in evaluating labour market and social inclusion policies. Use of counterfactual impact evaluation* (hierin worden 39 contrafeitelijke effectonderzoeken voor het ESF in de programmeringsperiode 2007-2013 geïdentificeerd), doi:10.2788/083390, <https://op.europa.eu/en/publication-detail/-/publication/d0e73217-2a40-47af-a5c3-a700dcd47da3/language-en>
- Bredgaard, T. (2015), "Evaluating What Works for Whom in Active Labour Market Policies", *European Journal of Social Security*, 17(4), blz. 436-452, <https://doi.org/10.1177/138826271501700403>
- Bryson, A., Dorsett, R. en Purdon, S. (2002), *The use of propensity score matching in the evaluation of active labour market policies*, Department for Work and Pensions, Working Paper Number 4, [http://eprints.lse.ac.uk/4993/1/The\\_use\\_of\\_propensity\\_score\\_matching\\_in\\_the\\_evaluation\\_of\\_active\\_labour\\_market\\_policies.pdf](http://eprints.lse.ac.uk/4993/1/The_use_of_propensity_score_matching_in_the_evaluation_of_active_labour_market_policies.pdf)
- Caliendo M., Kopeinig S. (2008), "Some practical guidance for the implementation of propensity score matching", *Journal of Economic Surveys*, Volume 22, Issue 1 — blz. 31-72, <https://doi.org/10.1111/j.1467-6419.2007.00527.x>
- Caliendo, M., Mahlstedt, R., & Mitnik, O. A. (2017), "Unobservable, but unimportant? The relevance of usually unobserved variables for the evaluation of labour market policies", *Labour Economics*, 46, blz. 14-25, <https://doi.org/10.1016/j.labeco.2017.02.001>
- Card, D., Ibarrran, P. en Villa, J. M. (2011), *Building in an evaluation component for active labour market programs: a practitioner's guide*, Discussion Paper No. 6085, Bonn, Duitsland: IZA, <http://ftp.iza.org/dp6085.pdf>
- Card D., Kluve J., Weber A. (2017), "What works? A meta-analysis of recent active labor market program evaluations", *Journal of the European Economic Association*, Volume 16, Issue 3, juni 2018, blz. 894–931, <https://doi.org/10.1093/jeea/jvx028>
- Centre for Disease Control and Prevention CDC (2013), *Good Evaluation Questions: Checklist to Help Focus Your Evaluation*, Department of Health & Human Services — USA, [http://www.cdc.gov/asthma/program\\_eval/AssessingEvaluationQuestionChecklist.pdf](http://www.cdc.gov/asthma/program_eval/AssessingEvaluationQuestionChecklist.pdf)
- Danmarks Statistik (2017), *Effektmåling og monitorering 2016 — strukturfondsindsatsen i Syddanmark*, <http://files.evaluationhelpdesk.eu/Evaluations/DKE7.pdf>
- Danmarks Statistik (2018), *Effektmåling af den Virksomhedsrettede Strukturfondsindsats 2007-2013*, <http://files.evaluationhelpdesk.eu/Evaluations/DKE10.pdf>
- Diaz J. M. et al. (2019), *Evaluación del PO-FSE 2014/2020 del Principado de Asturias para el informe anual a presentar en 2019*, <http://files.evaluationhelpdesk.eu/Evaluations/ESE101.pdf>
- Ecorys, Ismeri Europa (2020), *Study supporting the evaluation of ESF support to education and training (Thematic Objective 10) for the Commission, Directorate-General for Employment, Social Affairs and Inclusion*, <https://op.europa.eu/en/publication-detail/-/publication/d0c1a558-077d-11eb-a511-01aa75ed71a1/language-en/format-PDF/source-173162502>
- Europese Commissie (2007), *Indicative Guidelines on Evaluation Methods: Evaluation during the Programming Period*, Working paper no.5, DG Regionaal beleid, [https://ec.europa.eu/regional\\_policy/sources/docoffic/2007/working/wd5\\_ongoing\\_en.pdf](https://ec.europa.eu/regional_policy/sources/docoffic/2007/working/wd5_ongoing_en.pdf)
- Europese Commissie (2013), *EVALSED: The resource for the evaluation of Socio-Economic Development*, Updated version, [https://ec.europa.eu/regional\\_policy/sources/docgener/evaluation/guide/guide\\_evalsed.pdf](https://ec.europa.eu/regional_policy/sources/docgener/evaluation/guide/guide_evalsed.pdf)
- Europese Commissie (2017), *Toolbox voor betere regelgeving, Tool #47. Evaluation Criteria and*

Questions, [https://ec.europa.eu/info/sites/default/files/file\\_import/better-regulation-toolbox-47\\_en\\_0.pdf](https://ec.europa.eu/info/sites/default/files/file_import/better-regulation-toolbox-47_en_0.pdf)

Europese Commissie (2018), *Monitoring and Evaluation of European Cohesion Policy European Social Fund. Guidance document*, <https://ec.europa.eu/sfc/en/system/files/ged/ESF%20monitoring%20and%20evaluation%20guidance.pdf>

Europese Commissie (2019), *Advanced counterfactual evaluation methods. Guidance document*, doi:10.2767/464242, <https://ec.europa.eu/social/main.jsp?catId=738&langId=en&pubId=8254&furtherPubs=yes>

Europese Commissie (2020), *Counterfactual impact evaluation of European Social Fund interventions in practice. Guidance document for Managing Authorities*, doi:10.2767/55495, <https://op.europa.eu/en/publication-detail/-/publication/f82c5fb8-bb40-11ea-811c-01aa75ed71a1/language-en>

Europese Commissie (2020), *How to use administrative data for European Social Funds counterfactual impact evaluations. A step-by-step guide for managing authorities*, <https://op.europa.eu/en/publication-detail/-/publication/d96feed3-f30c-11ea-991b-01aa75ed71a1>

Verordening (EU) 2021/1060 van het Europees Parlement en de Raad van 24 juni 2021 houdende gemeenschappelijke bepalingen inzake het Europees Fonds voor regionale ontwikkeling, het Europees Sociaal Fonds Plus, het Cohesiefonds, het Fonds voor een rechtvaardige transitie en het Europees Fonds voor maritieme zaken, visserij en aquacultuur en de financiële regels voor die fondsen en voor het Fonds voor asiel, migratie en integratie, het Fonds voor interne veiligheid en het Instrument voor financiële steun voor grensbeheer en visumbeleid.

Europese Commissie (2021a): *Common Indicators Toolbox. Working Document*, juni 2021, [https://ec.europa.eu/sfc/en/system/files/2021/ged/toolbox\\_june\\_2021\\_final.pdf](https://ec.europa.eu/sfc/en/system/files/2021/ged/toolbox_june_2021_final.pdf)

Fondazione G. Brodolini, Metis GmbH, Applica, Ockham IPS (2020), *Study for the Evaluation of ESF Support to Employment and Labour Mobility (Thematic Objective 8), for the European Commission, Directorate-General for Employment, Social Affairs and Inclusion*, <https://ec.europa.eu/social/BlobServlet?docId=22899&langId=en>

Frolich, M. (2004), "Programme evaluation with multiple treatments", *Journal of Economic Surveys*, 18(2), blz. 181-224, <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.620.9209&rep=rep1&type=pdf>

Guthmuller S., Meroni E.C., *Evaluation of the higher education grant system for less privileged students in Portugal. JRC technical Report*. Nog niet gepubliceerd.

Hagglund, P. (2006), *A description of three randomised experiments in Swedish labour market policy*, Institute for Labour Market Policy Evaluation, Report 2006:4, <https://www.ifau.se/globalassets/pdf/se/2006/r06-04.pdf>

Heckman, J. J., Ichimura, H., Smith, J. and Todd, P. (1998), "Characterizing selection bias using experimental data", *Econometrica*, Vol. 66, No. 5, blz. 1017-1098, [http://jenni.uchicago.edu/papers/Heckman\\_Ichimura\\_etal\\_1998\\_Econometrica\\_v66\\_n5\\_r.pdf](http://jenni.uchicago.edu/papers/Heckman_Ichimura_etal_1998_Econometrica_v66_n5_r.pdf)

Hunger K. en Sattler K. (2017), *Evaluationsbericht zum spezifischen Ziel A 1.1. im Rahmen der Evaluierung des Operationellen Programms des Europäischen Sozialfonds in Baden-Württemberg 2014-2020*, <http://files.evaluationhelpdesk.eu/Evaluations/DEE26.pdf>

Holland, P. W. (1986), "Statistics and Causal Inference", *Journal of the American Statistical Association*, 81 (396), blz. 945-960, DOI: 10.1080/01621459.1986.10478354, <http://people.umass.edu/~stanek/pdf/cause/causal-holland.pdf>

IAW Institut für Angewandte Wirtschaftsforschung \_ ISG Institut für Sozialforschung und

- Gesellschaftspolitik, GmbH (2015), *Evaluation der Modellprojekte "Bürgerarbeit". Endbericht*, <http://files.evaluationhelpdesk.eu/Evaluations/DEE7.pdf>
- ICF, Cambridge Econometrics and Eurocentre (2020), *Study supporting the 2020 evaluation of promoting social inclusion, combating poverty and any discrimination by the European Social Fund (Thematic Objective 9) for the European Commission, Directorate-General for Employment, Social Affairs and Inclusion*, <https://op.europa.eu/en/publication-detail/-/publication/8788ec85-2308-11eb-b57e-01aa75ed71a1>
- Indecon (2016), *Indecon Evaluation of JobBridge Activation Programme*, <http://files.evaluationhelpdesk.eu/Evaluations/IEE2.pdf>
- Instytut Badań Strukturalnych, Imapp, IQS (2015), *Badanie efektów wsparcia zrealizowanego na rzecz osób młodych w ramach Programu Operacyjnego Wiedza Edukacja Rozwój. I Raport Tematyczny*, <http://files.evaluationhelpdesk.eu/Evaluations/PLE34.pdf>
- Ires Piemonte, (2019), *Misure di sostegno all'imprenditorialità*, <http://files.evaluationhelpdesk.eu/Evaluations/ITE61.pdf>
- Isfol (2016), *Primo rapporto di valutazione del piano italiano Garanzia giovani*, <http://files.evaluationhelpdesk.eu/Evaluations/ITE13.pdf>
- Ismeri Europa (2018), *I Rapporto tematico di valutazione – I risultati di PIPOL*, <http://files.evaluationhelpdesk.eu/Evaluations/ITE36.pdf>
- Ismeri Europa – Ecorys – Institute for Employment Studies (2019), *Pilot and feasibility study on the sustainability and effectiveness of results for European Social Fund participants using counterfactual impact evaluations*, doi:10.2767/39339, <https://op.europa.eu/it/publication-detail/-/publication/84cc9eb9-b33d-11e9-9d01-01aa75ed71a1>
- Ivanov B. et al., 2020, "Do job creation schemes improve the social integration and well-being of the long-term unemployed?", *Labour Economics*, <https://doi.org/10.1016/j.labeco.2020.101836>,
- Kalinowski H. (2020), *Badanie efektów wsparcia zrealizowanego na rzecz osób młodych w ramach Programu Operacyjnego Wiedza Edukacja Rozwój V Raport wskaźnikowy*, <http://files.evaluationhelpdesk.eu/Evaluations/PLE296.pdf>
- Kalinowski H. et al., 2020, *Badanie efektów wsparcia zrealizowanego na rzecz osób młodych w ramach Programu Operacyjnego Wiedza Edukacja Rozwój Raport końcowy*, <http://files.evaluationhelpdesk.eu/Evaluations/PLE339.pdf>
- Kluve, J., Lehmann, H., & Schmidt, C. M. (2008), "Disentangling Treatment Effects of Active Labor Market Policies: The Role of Labor Force Status Sequences", *Labour Economics*, 15(6), blz. 1270-1295, <https://doi.org/10.1016/j.labeco.2007.12.002>
- Krug, G en Stephan, G. (2011), *Is contracting-out intensified placement services more effective than in-house production? Evidence from a randomized field experiment*, LASER Discussion Papers — Paper No. 5, <http://doku.iab.de/externe/2011/k110912303.pdf>
- Lammers M. en Kok L. (2021), "Are active labor market policies (cost-)effective in the long run? Evidence from the Netherlands", *Empirical Economics* 60, blz.1719–1746, <https://doi.org/10.1007/s00181-019-01812-3>
- Lang, J. (2021), "Employment effects of language training for unemployed immigrants", *Journal of Population Economics*, 35, blz. 719–754, <https://doi.org/10.1007/s00148-021-00832-7>
- Martini, A. (2009), *Counterfactual impact evaluation: what it can (and cannot) do for cohesion policy, prepared for the 6th European Conference on Evaluation of Cohesion Policy, Warsaw, November 30th*, [https://ec.europa.eu/regional\\_policy/archive/conferences/evaluation2009/abstracts/martini.doc](https://ec.europa.eu/regional_policy/archive/conferences/evaluation2009/abstracts/martini.doc)
- Morgan, S. L. en Winship, C. (2014), *Counterfactual and causal inference: Methods and*

- principles for social research*, 2nd edition. Cambridge en New York: Cambridge University Press.
- OESO (2010), *Glossary of Key Terms in Evaluation and Results Based Management*, Parijs, <https://www.oecd.org/dac/evaluation/2754804.pdf>
- Openfield (2019), *Analiza skuteczności i efektywności dotacji na założenie działalności gospodarczej udzielonych w ramach 8 osi priorytetowej RPO WM — komponent 3*, <http://files.evaluationhelpdesk.eu/Evaluations/PLE312.pdf>
- Palczyńska M. et al. (2019), *Badanie efektów wsparcia zrealizowanego na rzecz osób młodych w ramach Programu Operacyjnego Wiedza Edukacja Rozwój, IV Raport wskaźnikowy*, <http://files.evaluationhelpdesk.eu/Evaluations/PLE274.pdf>
- Pomatto G. (2017), *L'attuazione del Buono per Servizi al Lavoro nella Regione Piemonte*, Ires Piemonte, <http://files.evaluationhelpdesk.eu/evaluations/ITE55.pdf>
- Pomatto G. (2019), *Buoni per servizi al lavoro nella Regione Piemonte: qualità percepita dai destinatari e meccanismi dell'attuazione*, Ires Piemonte, <http://files.evaluationhelpdesk.eu/evaluations/ITE77.pdf>
- Pompili M., Giorgetti I. (2020), *Rapporto di Placement. Servizio di attività di valutazione del POR FSE 2014/2020 — Regione Marche*, <http://files.evaluationhelpdesk.eu/Evaluations/ITE115.pdf>
- Pompili M., Giorgetti I. (2020a), *Rapporto tematico "Disoccupazione di lunga durata". Servizio di attività di valutazione del POR FSE 2014/2020 — Regione Marche*, <http://files.evaluationhelpdesk.eu/Evaluations/ITE214.pdf>
- Poy S. (2019), *Gli effetti occupazionali del buono per servizi al lavoro nella Regione Piemonte: prime evidenze. Misura per disoccupati da almeno 6 mesi, anno 2017*, Ires Piemonte, <http://files.evaluationhelpdesk.eu/evaluations/ITE123.pdf>
- Poy S. (2020), *Nuove evidenze sull'effetto occupazionale del buono per servizi al lavoro. Target persone disoccupate da almeno 6 mesi*, Ires Piemonte, <http://files.evaluationhelpdesk.eu/evaluations/ITE245.pdf>
- Re-source Pracownia Badań i Doradztwa (2015), *Ocena działań na rzecz wydłużenia aktywności zawodowej osób starszych w kontekście niekorzystnej sytuacji demograficznej w województwie lubelskim*, <http://files.evaluationhelpdesk.eu/evaluations/PLE126.pdf>
- Riccio J., Friedlander, D., Freedman S. (1994), *GAIN: Benefits, Costs, and Three-Year Impacts of a Welfare-to-Work Program*, MDRC, NYC <https://www.mdrc.org/publication/gain-benefits-costs-and-three-year-impacts-welfare-work-program>
- Scheller F. en Seidel K. (2020), *Zweiter Evaluationsbericht zum spezifischen Ziel A 1.1: Teilnehmer\*innenperspektive und Wirkungsanalyse. im Rahmen der Evaluierung des Operationellen Programms des Europäischen Sozialfonds in Baden-Württemberg 2014–2020*, <http://files.evaluationhelpdesk.eu/Evaluations/DEE171.pdf>
- Shadish, W. R., Cook, T. D. en Campbell, D. T. (2002), *Experimental and quasi-experimental designs for generalised causal inference*, Boston, US: Houghton Mifflin Company.
- Stern E. et al. (2012), *Broadening the range of designs and methods for impact evaluations*, Department for International Development of the UK, Working Paper 38, <https://assets.publishing.service.gov.uk/media/57a08a6740f0b6497400059e/DFIDWorkingPaper38.pdf>
- WK Kellogg Foundation (2004), *Logic Model Development Guide*, <https://www.wkkf.org/resource-directory/resources/2004/01/logic-model-development-guide>
- Zub M. et al. (2020), *Ocena wpływu wsparcia RPOWP na popularyzację szkolnictwa zawodowego w województwie podlaskim. Raport końcowy*, <http://files.evaluationhelpdesk.eu/Evaluations/PLE359.pdf>

# Bijlagen

## Bijlage 1. Literatuur

Hieronder volgt verdere literatuur voor personeelsleden van de beheerautoriteit die meer willen weten over de onderwerpen die in deze leidraad aan de orde komen. Er is uitgebreide literatuur over evaluaties voorhanden. Deze lijst is bedoeld om betrouwbare belangrijke behandelingen aan te geven die nuttige informatie voor CIE-planning opleveren. Na iedere verwijzing wordt van de meeste bronnen een beknopte beschrijving gegeven.

### Algemene evaluatie

- Rossi, Peter H., Mark W. Lipsey, en Gary T. Henry (2018), *Evaluation: A Systematic Approach*, 8th edition. Thousand Oaks, CA: SAGE Publications.

Het klassieke leerboek over evaluatiepraktijk en -methoden. Bevat methoden en voorbeelden.

- Gertler, Paul J., Martinez, Sebastian, Premand, Patrick, Rawlings, Laura B., Vermeersch, Christel M. J. (2016), *Impact Evaluation in Practice*, Second Edition. Washington, DC: Inter-American Development Bank and World Bank, © World Bank, <https://openknowledge.worldbank.org/handle/10986/25030> (Beschikbaar in het Engels, Portugees en Spaans).

Net als deze leidraad begint dit handboek met de klassieke (RCT) evaluatie om daarna in te gaan op alternatieve methoden. Hoewel deze leidraad is geschreven voor programmabeheerders in landen met lagere inkomens, is de discussie relevant en eenvoudig toepasbaar in de context van de EU-lidstaten.

- White, H., en Raitzer, D. A. (2017), *Impact evaluation of development interventions: A practical guide*, Asian Development Bank, <https://www.adb.org/sites/default/files/publication/392376/impact-evaluation-development-interventions-guide.pdf>

Net als de hierboven genoemde publicatie van Gertler et al. (2016) bevat deze gids — hoewel hij ingaat op het effectonderzoek van interventies in lage- en middeninkomenslanden — een uitgebreide behandeling van algemene CIE-methoden. Technischer dan Gertler et al. (2016), maar niettemin met veel praktische inzichten die interessant kunnen zijn voor ESF-programmamanagers.

- Europese Commissie (2019), *Advanced counterfactual evaluation methods. Guidance document*, doi:10.2767/464242, <https://op.europa.eu/en/publication-detail/-/publication/11968bbb-fac9-11e9-8c1f-01aa75ed71a1/language-en>

In dit document worden recente en meer geavanceerde methoden voor contrafeitelijk effectonderzoek gepresenteerd, zoals sequentieanalyse, dynamische matching en synthetische controles.

- Csillag Marton, Kreko Judi en Scharle Agota (2020), *Counterfactual evaluation of youth employment policies*, opgesteld in het kader van het “Youth Employment PartnerSHIP”-project, <http://yepartnership.ibs.org.pl/content/uploads/2021/02/Methodological-guide.pdf> (Beschikbaar in het Engels, Spaans, Hongaars (vertaling in uitvoering), Italiaans en Pools).

Dit is een “stapsgewijze” inleiding tot het contrafeitelijk onderzoek van arbeidsmarktbeleid voor jongeren, met de nadruk op het gebruik van administratieve gegevens. Vraagstukken worden gepresenteerd aan de hand van praktische problemen die zich hebben voorgedaan bij vier CIE's van subsidies voor het in dienst nemen van jongeren in Spanje, Hongarije, Italië

en Polen.

- HM Treasury (United Kingdom) (2020), *The Magenta Book: Guidance for evaluation*, Londen: The Agency, <https://www.gov.uk/government/publications/the-magenta-book>

In het “Magenta”-boek wordt uitvoerig ingegaan op evaluatiemethoden. Deze documenten zijn interessant als voorbeelden vanuit het oogpunt van interne evaluatie door de overheid.

### Gerandomiseerd onderzoek met een controlegroep (RCT)

- White H. (2013), “An introduction to the use of randomised control trials to evaluate development interventions”, *Journal of Development Effectiveness*, 5:1, blz. 30-49, DOI: 10.1080/19439342.2013.764652  
<https://www.tandfonline.com/doi/pdf/10.1080/19439342.2013.764652>

- White, H., Sabarwal S. & T. de Hoop, (2014), *Randomized Controlled Trials (RCTs), Methodological Briefs: Impact Evaluation 7*, UNICEF Office of Research, Florence, [https://www.unicef-irc.org/publications/pdf/brief\\_7\\_randomized\\_controlled\\_trials\\_eng.pdf](https://www.unicef-irc.org/publications/pdf/brief_7_randomized_controlled_trials_eng.pdf)

Twee niet-technische inleidingen in de logica van RCT, met een bespreking van verschillende opzetten en kritiek op RCT.

- Glennerster, R., en Takavarasha, K., (2013), *Running Randomized Evaluations: A Practical Guide*, Princeton University Press, Princeton, NJ.

Het boek is een stapsgewijze gids over hoe RCT's op het gebied van sociale programma's kunnen worden opgezet en uitgevoerd. Het is gebaseerd op de concrete RCT's die zijn uitgevoerd door Abdul Latif Jameel Poverty Action Lab.

### Difference-in-differences

- Card, David, Pablo Ibararán, en Juan Miguel Villa (2011), *Building in an Evaluation Component for Active Labor Market Programs: A Practitioner's Guide*, IZA Discussion Paper No. 6085. Bonn: Forschungsinstitut zur Zukunft der Arbeit, <http://ftp.iza.org/dp6085.pdf>

Vergelijkt DID met RCT.

- Lechner, M. (2011), “The Estimation of Causal Effects by Difference-in-Difference Methods”, *Foundations and Trends in Econometrics*, Vol. 4, No. 3 (2010), blz. 165–224, DOI: 10.1561/08000000014 [https://michael-lechner.eu/ml\\_pdf/journals/2011\\_Lechner\\_DiD\\_2011\\_ECO%200403%20Lechner\\_darf%20aufs%20Netz.pdf](https://michael-lechner.eu/ml_pdf/journals/2011_Lechner_DiD_2011_ECO%200403%20Lechner_darf%20aufs%20Netz.pdf)

In dit document wordt dieper ingegaan op de DID-benadering, evenals op enkele van de belangrijkste problemen bij de toepassing ervan. Uitbreidingen van DID, zoals niet-lineaire toepassingen en propensity score matching met DID, komen ook aan de orde.

- Fredriksson A. en de Oliveira G. M. (2019), “Impact evaluation using Difference-in-Differences”, *RAUSP Management Journal*, Vol. 54, No. 4, blz. 519-532, DOI 10.1108/RAUSP-05-2019-0112  
<https://www.emerald.com/insight/content/doi/10.1108/RAUSP-05-2019-0112/full/pdf?title=impact-evaluation-using-difference-in-differences>

Een overzicht van de DID-methoden, met praktische aanbevelingen.

- Card, David en Alan B. Krueger (1994), “Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania”, *American Economic Review*, 84 (4), blz. 772–793, <https://davidcard.berkeley.edu/papers/njmin-aer.pdf>

Het klassieke voorbeeld van het gebruik van een verschil-in-verschillen-methode.

### Instrumentele variabelen

- Morgan, Stephen L., en Christopher Winship (2014), *Counterfactuals and Causal Inference: Methods and Principles for Social Research*, 2nd edition. Cambridge en New York: Cambridge University Press.

Dit is een enigszins technisch overzicht van CIE-methoden waarbij gebruik wordt gemaakt van sociologische terminologie. Hoofdstuk 9, "Instrumental Variable Estimators of Causal Effects" (blz. 291-324) geeft een overzicht van de logica van en de procedures voor IV-schattingen.

- Kuhn, Andreas, Jean-Philippe Wuellrich en Josef Zweimüller (2010), *Fatal Attraction? Access to Early Retirement and Mortality*, IZA Discussion Paper No.5160, Bonn: Forschungsinstitut zur Zukunft der Arbeit, <http://ftp.iza.org/dp5160.pdf>

Gebruikt regionale variatie in de verandering van de pensioenleeftijd in Oostenrijk als instrumentele variabele in een onderzoek naar het effect van vervroegde pensionering op de gezondheid van werknemers.

- Galiani, S., Rossi, M. A., en Schargrodsy, E. (2011), "Conscription and crime: Evidence from the Argentine draft lottery", *American Economic Journal: Applied Economics*, <https://www.aeaweb.org/articles?id=10.1257/app.3.2.119>

Innovatief en illustratief gebruik van een gerandomiseerde loting (voor de dienstplicht) als instrumentele variabele. Zeer leesbare studie, prototypisch voor een onderzoekslijn die loterijprocedures als instrument gebruikt.

### Matching

- Heinrich, Carolyn, Alessandro Maffioli en Gonzalo Vázquez (2010), *A Primer for Applying Propensity Score Matching*, Impact-Evaluation Guidelines Technical Notes No. IDB-TN-161, Washington: Inter-American Development Bank, <https://publications.iadb.org/publications/english/document/A-Primer-for-Applying-Propensity-Score-Matching.pdf>

Net als de gids voor regressie-discontinuïteit hieronder, is dit document gericht op goed geïnformeerde evaluatiebeheerders.

- Caliendo M., Kopeinig S., (2008), "Some Practical Guidance for the Implementation of Propensity Score Matching", *Journal of Economic Surveys*, Volume 22, Issue 1 — blz. 31-72, <https://doi.org/10.1111/j.1467-6419.2007.00527.x>

Een klassieke tekst over de keuzes die moeten worden gemaakt bij de uitvoering van PSM (in termen van schatting, matching-algoritmen, kwaliteitsbeoordeling van de matching, gevoeligheid van geschatte behandelingseffecten).

- Harris, H. en Horst, S. J. (2016), "A Brief Guide to Decisions at Each Step of the Propensity Score Matching", *Process, Practical Assessment, Research, and Evaluation*, Vol. 21, Article 4, DOI: <https://doi.org/10.7275/yq7r-4820>. Beschikbaar op: <https://scholarworks.umass.edu/pare/vol21/iss1/4>

Vergelijkbaar met het vorige artikel.

### Regressie-discontinuïteitsanalyse

- Jacob, Robin, Pei Zhu, Marie-Andrée Somers, en Howard Bloom (2012), *A Practical Guide to Regression Discontinuity*, New York: MDRC, [https://www.mdrc.org/sites/default/files/RDD%20Guide\\_Full%20rev%202016\\_0.pdf](https://www.mdrc.org/sites/default/files/RDD%20Guide_Full%20rev%202016_0.pdf)

Buitengewoon toegankelijke en grondige bespreking van recessie-discontinuïteitsmethoden, met een zorgvuldig samengestelde bibliografie.

- Lee, D. S. en Lemieux T. (2009), *Regression discontinuity designs in economics*, NBER

Working Paper No. 14723, National Bureau of Economic Research, Cambridge, MA, 2009,  
[http://www.nber.org/papers/w14723.pdf?new\\_window=1](http://www.nber.org/papers/w14723.pdf?new_window=1)

Het document is een soort "handleiding" voor regressie-discontinuïtet (RD) dat een beschrijving geeft van de logica van de methode en toelichting verschaft bij de verschillende wijzen van het schatten van RD-analyses.



## Bijlage 2. Voorgestelde opzet voor een CIE-cursus

Een inleidende cursus over CIE's voor beheerautoriteiten en ambtenaren zou het volgende kunnen omvatten:

- Inleiding tot evaluatiemethoden: procesevaluaties (waarom zij belangrijk zijn en hoe ze verband houden met CIE's) en effectonderzoeken (contrafeitelijke en theoriegebaseerde methoden).
- Wat zijn CIE's? Op welke evaluatievragen kunnen krachtige CIE's een antwoord geven? Voor welke evaluatiecriteria leveren CIE's het belangrijkste bewijs? Waarom zijn CIE's belangrijk?
- Hoe werken ze? (Behandelde groep en controlegroep, "vóór en na"-vergelijkingen, inferentiebeginselen).
- Overzicht van CIE-methoden (kenmerken, sterke en zwakte punten van elke methode):
  - o gerandomiseerd onderzoek met een controlegroep,
  - o matching,
  - o verschil-in-verschillen,
  - o regressie-discontinuïteitsanalyse,
  - o instrumentele variabele,
  - o vermelding van andere mogelijke methoden (bv. technieken voor ruimtelijke analyses).
- Gegevensvereisten
  - o vereisten voor de behandelde groep en de controlegroep,
  - o mogelijke bronnen (ESF-monitoring, administratieve datasets, enquêtes),
  - o AVG-voorschriften.
- Overzicht van de uitvoeringsfasen:
  - o planning van CIE's (evaluatieplan, haalbaarheid van een CIE),
  - o opdracht geven tot CIE's (taakomschrijving, selectiecriteria en methoden),
  - o beheer van CIE's (controle van de te leveren prestaties en interactie met de beoordelaar),
  - o verspreiding van de bevindingen van CIE's (soorten publiek en communicatiekanalen).

Een cursus met deze opzet moet praktische voorbeelden bevatten die in workshops worden uitgewerkt en duurt 2-3 dagen. Een benadering voor het geven van een cursus als deze zou erin bestaan een methode van probleemgestuurd onderwijs te hanteren en een beleidsmaatregel van het programma te gebruiken als het concrete geval dat moet worden ontwikkeld.

Bijlage 3 Contrafeitelijke effectonderzoeken — in de leidraad genoemde voorbeelden

Tabel 6 Kenmerken van de CIE's die in de leidraad als voorbeeld worden genoemd

Land	Titel — jaar	Geëvalueerd beleid en geobserveerde behandelingsperiode	Methode	Behandelde groep en controlegroep — vaststelling en omvang	Uitkomstvariabelen	Observatieperiode	Gegevens
Zweden	Effecten van intensievere ondersteuning voor jonge werkzoekenden — Resultaten van de evaluatie van "Jonge toekomst" (2019)	Intensievere ondersteuning van werkzoekenden door de openbare dienst voor arbeidsvoorziening Juni 2017 — januari 2018	Gerandomiseerd onderzoek met een controlegroep (RCT)	BG: 18- tot 24-jarigen die willekeurig zijn toegewezen om intensievere steun te ontvangen CG: 18- tot 24-jarigen die willekeurig zijn toegewezen om normale steun te ontvangen BG: 2 792 — CG: 1 897	- aandeel werklozen - % uitstroom naar werk - gemiddeld aantal dagen zonder werk	12 maanden na toetreding	- openbare werkgelegenheidsregisters voor uitkomstvariabelen - enquête onder arbeidsbemiddelaars om de intensiteit van de verleende steun te meten
Italië (Marche)*	De effecten van ESF-interventies die in 2014-2020 zijn gefinancierd voor langdurig werklozen in de regio Marche (2020)	Stages, werkervaringen, werkbeurzen, opleidingsvouchers voor werklozen (waaronder langdurig werklozen) 2017 — juli 2019	Propensity score matching (PSM)	BG: behandelde langdurig werklozen CG: bij ODA's geregistreerde langdurig werklozen in de periode 2016-2018 BG: 526 met stages, 1 058 met werkbeurzen, 236 met werkervaring in gemeenten en 241 met opleidingsvouchers — CG: 77 255	- waarschijnlijkheid van tewerkstelling op een bepaald tijdstip na de interventies - waarschijnlijkheid van tewerkstelling met een overeenkomst voor onbepaalde tijd - aantal gewerkte dagen in een periode na de interventies	6, 9, 12, 15 en 18 maanden na toetreding tot de interventies	- ESF-monitoringgegevens voor behandelde groep - administratieve gegevens uit ODA's over geregistreerde personen - administratieve gegevens over nieuwe arbeidsovereenkomsten in de regio ( <i>Comunicazioni Obbligatorie</i> — COB)
Polen (Lubelskie)*	Evaluatie (2015) van uit het ESF gefinancierde maatregelen ten behoeve van oudere werknemers in Lubelskie in de programmeringsperiode 2007-2013	Verschillende soorten actief arbeidsmarktbeleid 2007-2013	PSM + verschil-in-verschillen (DID)	BG: steekproef van werkloze 50-plussers die steun ontvingen CG: werkloze 50-plussers die aan geen enkele maatregel hebben deelgenomen en die op het moment dat de behandelde groepsleden deelnamen aan het geëvalueerde programma bij een regionaal arbeidsbureau waren ingeschreven BG: 145 — CG: 67 102	- waarschijnlijkheid om de toestand van geregistreerde werkloze te verlaten	6, 12, 18 maanden na toetreding tot de interventies	- administratieve gegevens van ODA's

Land	Titel — jaar	Geëvalueerd beleid en geobserveerde behandelingsperiode	Methode	Behandelde groep en controlegroep — vaststelling en omvang	Uitkomstvariabelen	Observatieperiode	Gegevens
Letland	Beroepsopleiding voor werkloze jongeren in Letland: bewijs afkomstig van een regressie-discontinuïteitsanalyse (2018)	Beroepsopleiding gericht op NEET's van 15-29 jaar 2014-2015	Regressie-discontinuïteitsanalyse (RDD)	BG: werklozen die binnen een jaar na de datum van registratie aan de beroepsopleiding hebben deelgenomen  CG: tussen juni 2013 en december 2015 geregistreerde werklozen die aan geen enkele maatregel deelnemen  BG: 898 — CG: 10 717	- waarschijnlijkheid van tewerkstelling op een bepaald tijdstip na de interventies - inkomen	12-36 maanden na inschrijving	- administratieve gegevens afkomstig van het Letse staatsbureau voor arbeidsvoorziening (SEA) — informatie over zowel deelnemers als niet-deelnemers die op specifieke data als werkloos zijn geregistreerd - administratieve gegevens afkomstig van de belastingdienst (SRS) — informatie over de arbeidsstatus op verschillende data en over het inkomen van de personen
Duitsland	Effecten van een taal cursus voor werkloze immigranten op arbeidsparticipatie (2021)	Taalopleidingen (+ werkervaringen voor sommige van de behandelde personen) voor migrerende werkzoekenden 2014	Instrumentele variabelen (IV)	BG: migrerende deelnemers die in 2014 zijn begonnen met een taalopleiding voor beroepsdoeleinden  CG: willekeurige steekproef van niet-deelnemers met ten minste één periode van niet-Duits staatsburgerschap die niet deelnemen aan actieve arbeidsmarktmaatregelen of integratiecursussen  BG: 8 968 — CG: 26 463	- waarschijnlijkheid van tewerkstelling (reguliere arbeid, reguliere voltijdse arbeid, reguliere arbeid > 6 maanden, arbeid met een inkomen boven de armoederisicodrempel, geschoolde arbeid) - dagelijks en gecumuleerd inkomen uit reguliere arbeid - gecumuleerde dagen in reguliere arbeid	24 maanden na toetreding tot de interventies	- geïntegreerde beroepsbiografieën (IEB) voor de identificatie van beide groepen en voor uitkomstvariabelen - gegevens die achteraf worden verzameld tijdens vergaderingen met werkzoekenden en betrokken dossierbeheerders (Werdegangshistorie, WGH), voor meer informatie over perioden van zelfstandige arbeid of ouderschapsverlof en over onderwijs en werk in het buitenland
Italië (Trento)	Zijn beroepsopleidingsprogramma's hun kosten waard? Empirische gegevens van een kosten-batenanalyse (2018)	Beroepsopleidingen voor werklozen 2010-2011	PSM	BG: werkloze deelnemers aan beroepsopleidingen, gefinancierd uit provinciale middelen en uit het ESF CG: personen die op de begindatum van de individuele opleidingen als werkloos zijn ingeschreven in de registers van de lokale ODA's  BG: 818 (provinciale cursussen), 114 (ESF-cursussen) — CG: 17 236 (verbonden aan provinciale cursussen), 1 152 (verbonden aan ESF-cursussen)	- waarschijnlijkheid van tewerkstelling - loon	3, 6, 12, 18, 24 en 36 maanden na toetreding tot de interventies	- ESF-monitoringgegevens voor behandelde groep - administratieve gegevens uit ODA's over geregistreerde personen - administratieve gegevens over nieuwe arbeidsovereenkomsten in de regio ( <i>Comunicazioni Obbligatorie</i> — COB) - administratieve gegevens over belastingaangiften van INPS - administratieve gegevens

Land	Titel — jaar	Geëvalueerd beleid en geobserveerde behandelingsperiode	Methode	Behandelde groep en controlegroep — vaststelling en omvang	Uitkomstvariabelen	Observatieperiode	Gegevens
							van het provinciaal ministerie van Onderwijs
Duitsland	Kunnen regelingen voor het scheppen van banen de sociale integratie en het welzijn van langdurig werklozen verbeteren? (2020)	Regeling voor het scheppen van banen (gesubsidieerde banen voor maximaal 36 maanden) voor kwetsbare personen (langdurig werklozen) 2015 — juni 2017	PSM	BG: steekproef van deelnemers die de volledige vragenlijst hebben beantwoord CG: bij ODA's geregistreerde niet-deelnemers met dezelfde criteria voor het al dan niet in aanmerking komen als de behandelde groep, die de volledige vragenlijst hebben beantwoord BG: 2 531 in de eerste ronde, 1 191 in de tweede ronde, 450 in de derde ronde — CG: zoals hierboven.	- waarschijnlijkheid van tewerkstelling - subjectieve maatregelen van persoonlijke tevredenheid, geestelijke gezondheid, maatschappelijke verbondenheid, sociale status	7, 18, 29 maanden vanaf toetreding tot het programma	- geïntegreerde beroepsbiografieën (IEB) voor de identificatie van beide groepen en voor uitkomstvariabelen - enquête voor subjectieve uitkomstvariabelen (soft outcomes)
Spanje (Asturië)*	Evaluatie van het ESF-OP van Asturië (2019)	Onderwijsmethoden voor groepen leerlingen met een risico van voortijdig schoolverlaten (tweede en derde jaar van de middelbare school) 2016-2017	PSM	BG: leerlingen die aan de interventies deelnemen CG: vergelijkbare leerlingen wat ontvankelijkheidscriteria betreft BG: 1 053 — CG: 3 852	- waarschijnlijkheid om over te gaan naar het volgende studiejaar	12 maanden na toetreding	- administratieve gegevens over onderwijs: SAUCE-database (ministerie van Onderwijs)
Portugal	Evaluatie van het stelsel van studiebeurzen voor het hoger onderwijs voor kansarme studenten in Portugal (2020) — te publiceren	Studiebeurzen voor het hoger onderwijs voor studenten uit gezinnen met een laag inkomen 2012-2018	RDD	BG: studenten die zich inschrijven voor het eerste jaar van een academische master of bachelor CG: studenten die vanwege een te hoog inkomen niet in aanmerking komen BG: 130 602 — CG: 25 400	- kans op inschrijving aan het einde van het eerste jaar - aantal behaalde studiepunten aan het einde van het eerste jaar - kans om af te studeren - aantal jaren om af te studeren - eindcijfer bij afstuderen	12-48 maanden na de studiebeurs	- administratieve gegevens van studenten die de beurs aanvragen (DGES) - administratieve gegevens met informatie over academische carrière en vooruitgang (DGEEC)
Polen (Podlaskie)*	Beoordeling van het effect op de popularisering van het ondersteunend beroepsonderwijs uit het OP Podlaskie in 2014-2020 (2020)	ESF-project "Goede baan — geweldig leven", ter bevordering van beroepsonderwijs en -opleiding en gericht op het lager secundair onderwijs 2017-2019	PSM	BG: steekproef van deelnemende leerlingen CG: niet aan het project deelnemende leerlingen met vergelijkbare kenmerken als die van de behandelde groep BG: 200 — CG: 384	- inschrijving bij een technische school op het moment van het onderzoek - inschrijving bij een eerstegraadsvakschool op het moment van het onderzoek - inschrijving bij een technische of een eerstegraadsvakschool op het moment van het onderzoek - aanmelding bij een middelbare school, technische school,	12-36 maanden na de interventies	- monitoringgegevens - nationaal bevolkingsregister (PESEL) voor selectie van de controlegroep - enquête voor meting van uitkomstvariabelen

Land	Titel — jaar	Geëvalueerd beleid en geobserveerde behandelingsperiode	Methode	Behandelde groep en controlegroep — vaststelling en omvang	Uitkomstvariabelen	Observatieperiode	Gegevens
					eerstegraadsvakschool, school voor beroepsopleiding		
Italië (Piemonte)*	Effecten op arbeidsparticipatie van de vouchers in het kader van de regeling inzake arbeidsvoorzieningsdiensten (Buono servizi lavoro — BSL, gefinancierd uit het ESF-OP Piemonte, 2014-2020) (2019)	Vouchers voor arbeidsvoorzieningsdiensten (oriëntatie + opleiding + stage) voor kwetsbare werklozen met een werkloosheidsperiode van ten minste 6 maanden 2018	PSM	BG: werkloze behandelde personen CG: bij ODA's geregistreerde werklozen met vergelijkbare kenmerken (30 jaar oud en 6 maanden werkloos) BG: 8 125 — CG: 130 000	- waarschijnlijkheid van tewerkstelling - waarschijnlijkheid van tewerkstelling met een overeenkomst voor onbepaalde tijd	6, 12, 16 maanden na de interventies	- ESF-monitoringgegevens voor behandelde groep - administratieve gegevens uit ODA's over geregistreerde personen - administratieve gegevens over nieuwe arbeidsovereenkomsten in de regio ( <i>Comunicazioni Obbligatorie</i> — COB)
Duitsland*	Evaluatie van het programma voor de integratie van langdurig werklozen in Duitsland in 2014-2020 (2019)	Integratiemaatregelen voor langdurig (24 maanden of langer) werklozen (gesubsidieerde banen) 2015-2017	DID (Intentie tot deelname was geschat)	BG: potentieel in aanmerking komende langdurig werklozen tijdens de duur van het programma (vanaf 2015) CG: personen die vóór de uitvoering van het programma aan de ontvankelijkheidscriteria voldoen (2010-2012) BG en CG: 134 515 personen voor 237 874 perioden van werkloosheid (35 % voor BG en 65 % voor CG)	- waarschijnlijkheid van tewerkstelling (tewerkstelling met betaling van sociale bijdragen)	24 maanden na de interventies	- administratieve monitoringgegevens van het Federale Bureau voor administratie (Zuwes-databank) - geïntegreerde beroepsbiografieën (IEB) voor uitkomstvariabelen
Duitsland (Baden-Württemberg)*	Tweede evaluatieverslag over de specifieke doelstelling A1.1 in het kader van het ESF-OP Baden-Württemberg, 2014-2020 (2020)	Maatregelen voor de integratie van werklozen, met name langdurig werklozen en andere kwetsbare personen Maart 2016-december 2017	PSM	BG: deelnemer geïdentificeerd in de IEB-databank CG: personen die in dezelfde periode aan de ontvankelijkheidscriteria voldoen BG: 1 578 (op een totaal van 1 800) — CG: n.v.t.	- waarschijnlijkheid van tewerkstelling (tewerkstelling met betaling van sociale bijdragen) - waarschijnlijkheid van aanhouding van een werkloosheidsuitkering	15 maanden na de interventies	- monitoringgegevens - geïntegreerde beroepsbiografieën (IEB) voor uitkomstvariabelen

Opmerking: BG = behandelde groep, CG=controlegroep.

Opmerking: \* verwijst naar in de Helpdesk geïdentificeerde evaluaties.

## CONTACT OPNEMEN MET DE EU

### Kom langs

Er zijn honderden Europe Direct-informatiecentra overal in de Europese Unie. U vindt het adres van het dichtstbijzijnde informatiecentrum op:

[https://europa.eu/european-union/contact\\_nl](https://europa.eu/european-union/contact_nl)

### Bel of mail

Europe Direct is een dienst die uw vragen over de Europese Unie beantwoordt. U kunt met deze dienst contact opnemen door:

- te bellen naar het gratis nummer: 00 800 6 7 8 9 10 11 (bepaalde telecomaandieners kunnen wel kosten in rekening brengen),
- te bellen naar het gewone nummer: +32 22999696, of
- een e-mail te sturen via: [https://europa.eu/european-union/contact\\_nl](https://europa.eu/european-union/contact_nl)

## WAAR VINDT U INFORMATIE OVER DE EU?

### Online

Informatie over de Europese Unie in alle officiële talen van de EU is beschikbaar op de Europa-website op: [https://europa.eu/european-union/index\\_nl](https://europa.eu/european-union/index_nl)

### EU-publicaties

U kunt publicaties van de EU downloaden of bestellen op: <https://op.europa.eu/nl/publications> Als u meerdere exemplaren van gratis publicaties wenst, neem dan contact op met Europe Direct of uw plaatselijke informatiecentrum (zie [https://europa.eu/european-union/contact\\_nl](https://europa.eu/european-union/contact_nl)).

### EU-wetgeving en aanverwante documenten

Toegang tot juridische informatie van de EU, waaronder alle EU-wetgeving sinds 1952 in alle officiële talen, krijgt u op EUR-Lex op: <http://eur-lex.europa.eu>

### Open data van de EU

Het opendataportaal van de EU (<http://data.europa.eu/euodp/nl>) biedt toegang tot datasets uit de EU. Deze gegevens kunnen gratis worden gedownload en hergebruikt, zowel voor commerciële als voor niet-commerciële doeleinden.



Publications Office  
of the European Union