



# WHITEPAPER

---

## WETEN OF HET WERKT - DE ROL VAN CONTROLEGROEPEN IN DE PRAKTIJK

Impact  
Centre  
Erasmus



# COLOFON

## Auteurs

David van de Velde  
Pauline Owelle

## Redactie

Marit Schouten

## Vormgeving

Charlotte Put

## Uitgave van

Impact Centre Erasmus, Rotterdam  
28 augustus 2025

## Dankwoord

Met dank aan alle collega's van het Impact Centre Erasmus:

- Karen Maas
- Jacqueline Scheidsbach
- Marjelle Vermeulen
- Paul Ongenae
- Emiel Gieles
- Ignace Breemer

# OVER

*Impact Centre Erasmus (ICE) is een onderwijs- en onderzoeksinstituut verbonden aan de economische faculteit van de Erasmus Universiteit Rotterdam. Impact Centre Erasmus wil met impact denken, meten en sturen organisaties stimuleren de grootst haalbare maatschappelijke bijdrage te leveren. Dit zodat mensen en middelen bewust en effectiever worden ingezet om economische, maatschappelijke en ecologische ambities te realiseren. Impact Centre Erasmus werkt met en voor het bedrijfsleven, de overheid, NGO's en goede doelen om hen te ondersteunen weloverwogen keuzes te maken.*

Impact Centre Erasmus viert dit jaar een jubileum! 20 jaar expertise en 10-jarig bestaan van het expertisecentrum voor onderzoek en onderwijs. Impact ligt nu overal op tafel. We hoeven niet steeds meer uit te leggen waarom impact ertoe doet. Ook niet dat we verder moeten kijken dan 'wat we hebben gedaan'. Het gaat juist om antwoorden op de vraag 'wat hebben we bereikt'? We zetten nu het pad in van impactmeten naar impactdenken. Ter ere van dit jubileum delen we onze kennis – met iedereen die het maar wilt lezen.

Ter ere van dit jubileum delen we onze kennis middels *whitepapers*. In een *whitepapers* nemen we een veelgehoord impactvraagstuk onder de loep. We delen de wetenschappelijke kennis hierover, en combineren dit met onze opgedane ervaring in hoe deze kennis in de praktijk ingezet kan worden. Zo zetten we het belang van impactdenken en impactsturen op de kaart én vergroten we samen de capaciteit van organisaties om meer impact te maken!

## INHOUDSOPGAVE

INHOUDSOPGAVE.....	3
1. HET BELANG VAN DE CONTROLEGROEP VOOR IMPACT.....	4
2. BELANGRIJKE DEFINITIES.....	6
3. DE CONTROLEGROEP.....	7
4. CONTROLEGROEP = INTERVENTIEGROEP.....	10
5. VERGELIJKING MET GROTERE GROEP.....	11
6. MATCHING OP GROEPSNIVEAU.....	12
7. MATCHING OP PERSOONSNIVEAU.....	14
8. RANDOMISATIE.....	16
9. CONCLUSIE EN AANBEVELINGEN.....	18
10. BRONVERMELDING.....	20

### Deze *whitepaper* gaat over methodes om controlegroepen op te zetten:

- ✓ Werk je met sociaal-maatschappelijke interventies waarbij je impact beoogt op een (groep) deelnemers?
- ✓ Zoek je naar manieren om het effect van jullie interventie aan te tonen?
- ✓ Wil jij weten wat een controlegroep is?
- ✓ En wanneer en hoe je een controlegroep kunt inzetten?

Dan is deze *whitepaper* voor jou!

# 1. HET BELANG VAN DE CONTROLEGROEP VOOR IMPACT

Impact Centre Erasmus definieert impact als het **extra effect** van een organisatie op stakeholders en de maatschappij, op de economische, sociaal-maatschappelijke en ecologische dimensies. Impact is in deze definitie een type effect, dat zowel positief en/of negatief, direct en/of indirect, bedoeld en/of onbedoeld kan zijn. Daarnaast is het een **extra** effect. Hiermee wordt bedoeld dat er pas over impact gesproken kan worden als er sprake is van een verschil ten opzichte van het effect dat mogelijk was opgetreden zonder interventie<sup>1</sup> van de organisatie – dat wat er sowieso zou zijn gebeurd.

Veel organisaties en initiatieven vinden het moeilijk om vast te stellen wat het **extra** effect van hun acties is, en of gevonden effecten ook zouden zijn opgetreden zonder hun inspanningen. Om te kunnen bepalen wat het **extra** effect is, oftewel het deel van een gevonden effect dat toe te wijzen is aan een interventie, is het nodig een vergelijking te maken met een groep die (nog) niet aan de interventie heeft deelgenomen. De resultaten van een impactmeting zijn immers altijd het resultaat van een vergelijking: het effect van de interventie ten opzichte van een situatie waarin de interventie niet plaatsvindt. Een veelvoorkomende manier om sociaal-maatschappelijke<sup>2</sup> effecten bij mensen vergelijkbaar te maken, is door te werken met een controlegroep. De controlegroep dient als vergelijking ten opzichte van de groep die wél aan een interventie deelneemt. Dit stelt in staat om te bepalen welk deel van het effect bij iedereen optreedt (deelnemers én controlegroep) en wat het **extra** effect van een interventie is (alleen deelnemers).

## *Voorbeeld: Bijlesprogramma*

Er wordt een bijlesprogramma gestart voor jongeren met als doel hun cijfers te verbeteren. Wanneer een jongere na deelname aan het programma hogere cijfers haalt zijn hiervoor verschillende verklaringen mogelijk. De hogere cijfers kunnen komen door het bijlesprogramma, maar het kan ook komen omdat docenten op de school van lesprogramma zijn gewisseld, of omdat de jongere naast het bijlesprogramma ook nog deelneemt aan een ander initiatief in de wijk. In dit geval is de bijdrage van het bijlesprogramma aan de hogere cijfers inzichtelijk te maken door ook resultaten van jongeren die wel en niet deelnemen aan het bijlesprogramma tegen elkaar af te zetten.

Deze *whitepaper* geeft inzicht in de mogelijkheden om controlegroepen samen te stellen en toont **wanneer** en **hoe** het inzetten van een controlegroep waardevol kan zijn om impact aan te tonen. Hierbij is een selectie gemaakt van vijf methoden om controlegroepen te vormen. Deze selectie is gemaakt op basis van hun relevantie voor de praktijk. De methoden variëren in complexiteit van de uitvoering en robuustheid van het bewijs. Deze methoden zijn relevant voor organisaties die op dit moment al interventies hebben met een sociaal-maatschappelijke focus<sup>3</sup> en die een eventueel gevonden effect bij hun deelnemers willen toewijzen aan hun interventie.

<sup>1</sup> In deze *whitepaper* wordt 'interventies' als koepelterm gebruikt voor verschillende typen inspanningen vanuit een organisatie, zoals activiteiten, programma's en projecten.

<sup>2</sup> Voor ecologische of structurele effecten (zoals veranderingen in biodiversiteit, luchtkwaliteit of stedelijke veranderingen) zijn controlegroepen minder gebruikelijk. In die gevallen wordt vaak gewerkt met andere vormen van vergelijking, zoals historische data, vergelijkbare gebieden (*matched comparisons*) of modelmatige schattingen van het verwachte effect zonder ingrijpen (*counterfactuals*).

<sup>3</sup> Voor ecologische initiatieven zijn andere vergelijkingsmethoden gangbaarder dan het gebruik maken van een controlegroep

### STRUCTUUR VAN DE *WHITEPAPER*

De *whitepaper* start met achtergrondinformatie over wat een controlegroep is en hoe en wanneer het gebruikt kan worden. Daarna worden vijf methoden uitgewerkt door middel van een korte uitleg, een visualisatie van de methode en een praktisch voorbeeld. Het praktische voorbeeld is steeds een bijlesprogramma voor jongeren. Door steeds hetzelfde voorbeeld te hanteren worden de verschillen tussen de methoden duidelijker.

Bij iedere methode is een tabel toegevoegd waarin vier categorieën worden beschreven:

- i. Complexiteit van uitvoerbaarheid
- ii. Moment van samenstellen controlegroep
- iii. Robuustheid van het bewijs
- iv. Ethisch dilemma

Vervolgens wordt de beschrijving van elke methode afgesloten met aandachtspunten en tips wanneer de methode gebruikt kan worden.

## 2. BELANGRIJKE DEFINITIES

Definities	Uitleg
Attributie	Betekent dat (een deel van) het effect kan worden toegeschreven aan een specifieke interventie, en niet veroorzaakt wordt door andere (externe) factoren. <sup>i</sup>
Bias <sup>4</sup>	Een (vaak onbewuste) voorkeur of vooroordeel waardoor mensen, data of ideeën niet eerlijk of objectief worden onderzocht. Bias kan ontstaan in hoe informatie wordt verzameld, geïnterpreteerd of gepresenteerd, en kan leiden tot onjuiste uitkomsten. <sup>ii</sup>
Causaliteit	Ook wel een oorzaak-gevolg relatie, waarbij wordt verondersteld dat een verandering in de ene variabele de oorzaak is van een verandering in de andere. <sup>iii</sup>
Controlegroep	Een groep die niet aan de interventie wordt blootgesteld om de effecten van een interventie te kunnen vergelijken met de interventiegroep. De vergelijking kan gemaakt worden tussen verschillende groepen, tussen dezelfde groep op verschillende momenten (T0-T1), of tussen groepen die verschillende interventies of omstandigheden meemaken. <sup>iv</sup>
Correlatie	Twee variabelen veranderen samen zonder dat de ene noodzakelijkerwijs de oorzaak is van de andere. <sup>v</sup>
Externe validiteit	Gaat over de mate waarin de conclusies van een onderzoek gegeneraliseerd kunnen worden naar de bredere populatie of andere situaties, hiervoor is het belangrijk dat de groep die onderzocht wordt representatief is voor de grotere groep waar het onderzoek uitspraken over wil maken. <sup>vi</sup>
Externe factoren	Andere factoren naast of buiten de interventie die mogelijk ook effect hebben/van invloed zijn op het resultaat. Dit worden ook wel <i>confounders</i> genoemd. <sup>vii</sup>
Impact	Het extra effect van een organisatie op de stakeholders en de maatschappij, op de economische, sociaal-maatschappelijke en ecologische dimensies. Impact is een type effect, dat zowel positief en/of negatief, direct en/of indirect, bedoeld en/of onbedoeld kan zijn.
Interne validiteit	Verwijst naar de mate waarin een vastgesteld effect in een onderzoek daadwerkelijk kan worden toegeschreven aan de onderzochte interventie en niet aan andere of externe factoren. <sup>viii</sup>
Randomisatie	Een proces waarbij personen willekeurig worden toegewezen aan een groep die wel, geen of een andere interventie krijgt. <sup>ix</sup>

<sup>4</sup> Er bestaan veel verschillende vormen van bias (zoals selectie- en informatie-bias), elk met specifieke oorzaken en gevolgen. In deze *whitepaper* wordt niet verder ingegaan op de verschillende typen bias, omdat dit buiten de scope valt.

## 3. DE CONTROLEGROEP

### 3.1 Wat is een controlegroep?

De controlegroep is een groep individuen, organisaties of een andere onderzoekspopulatie die *niet* deelnemen aan een interventie, maar verder zoveel mogelijk vergelijkbaar zijn met de groep die de interventie *wel* ontvangt (de interventiegroep<sup>x</sup>). Door de uitkomsten van beide groepen te vergelijken, kan worden vastgesteld of en zo ja, in welke mate de interventie effect heeft gehad.

Om betrouwbare conclusies te trekken, is het essentieel dat de controlegroep zoveel mogelijk overeenkomt met de interventiegroep, met idealiter het enige verschil dat de controlegroep niet de interventie heeft ontvangen<sup>xi</sup>. Dit vereist een zorgvuldige selectie van degenen in de controlegroep en voldoende data om de controlegroep te vergelijken met de interventiegroep. Het doel is om het effect van de interventie nauwkeurig te isoleren door te kijken naar de verschillen in uitkomsten tussen de groepen.

Al spreekt dit misschien wel voor zich, toch zijn er twee belangrijke punten te benoemen bij het samenstellen van controlegroepen. Allereerst moet een groep beschikbaar zijn. Het draait namelijk niet alleen om het identificeren van een vergelijkbare groep, maar ook om de toegang tot die groep. Dit is niet altijd vanzelfsprekend. Ten tweede is de grootte van de controlegroep van belang. Dit hangt af van verschillende factoren, zoals hoe groot het verwachte effect is, hoe sterk het bewijs moet zijn, en hoe groot en divers de totale populatie is. Als een sterk effect wordt verwacht vanuit een interventie kunnen soms minder mensen in de controlegroep volstaan. Maar hoe subtieler het verschil is dat aangetoond moet worden, hoe groter de groep moet zijn om betrouwbare conclusies te kunnen trekken.

#### Voorbeeld: Bijlesprogramma

Om te kunnen onderzoeken of het eerdergenoemde bijlesprogramma voor jongeren effect heeft op de cijfers van de deelnemers kan hier gebruik worden gemaakt van een controlegroep. De controlegroep is in dit voorbeeld een groep jongeren die geen bijles krijgt, maar bijvoorbeeld wel in hetzelfde schooljaar zitten en hetzelfde gemiddelde cijfer hebben als de groep jongeren die wel bijles krijgt. Wanneer het cijfer stijgt van de jongeren die bijles krijgen, maar het cijfer gaat niet omhoog van de jongeren die geen bijles krijgen (de controlegroep), is het aannemelijk dat de interventie (het krijgen van bijles) een positief effect heeft op het cijfer.

Het is in dit geval van belang dat de controlegroep overeenkomt met de groep jongeren die bijles krijgt (de interventiegroep). Dat betekent dat de jongeren in beide groepen bijvoorbeeld dezelfde leeftijd hebben, op hetzelfde opleidingsniveau zitten en eenzelfde gemiddeld cijfer hebben voor de vakken waarin zij bijles ontvangen. Hiermee wordt de kans groter dat het gevonden effect is toe te wijzen aan de bijles, in plaats van aan andere factoren.

Controlegroepen helpen om de causale relatie (oftewel de oorzaak-gevolg redenering) tussen een interventie en de gevonden effecten inzichtelijk te maken<sup>xii</sup>. Een effect dat in dezelfde mate bij zowel de controle- als de interventiegroep wordt gevonden, kan niet veroorzaakt zijn door de interventie aangezien de controlegroep die niet heeft ontvangen. Dit effect wordt dan veroorzaakt door externe factoren. Externe factoren, zoals economische ontwikkelingen, beleidswijzigingen, of persoonlijke omstandigheden, kunnen invloed hebben op de geobserveerde effecten.

### 3.2 Methodes om een controlegroep op te zetten.

In deze *whitepaper* wordt een aantal methoden besproken om een controlegroep op te zetten. De volgende methoden worden besproken:

1. Controlegroep = interventiegroep
2. Vergelijking met een grotere groep
3. *Matching* op groepsniveau
4. *Matching* op persoonsniveau
5. Randomisatie

Deze geselecteerde methoden hebben als doel om de causale effecten van interventies te isoleren, maar variëren in uitvoerbaarheid en de robuustheid van het bewijs dat de methode oplevert. De keuze voor een methode hangt af van de context waarin de interventie wordt uitgevoerd, de beschikbare middelen en de gewenste robuustheid van de resultaten.

De methoden om een controlegroep op te zetten worden in de volgende hoofdstukken uitgelegd. Daarnaast worden ze op aspecten met elkaar vergeleken. *Ten eerste* wordt gekeken naar de uitvoerbaarheid van een methode. Sommige methoden vereisen meer middelen zoals geld, tijd en kennis dan andere methoden. *Ten tweede* worden de methodes vergeleken ten opzichte van de robuustheid van het bewijs. Wanneer het bewijs robuuster is, is er meer zekerheid over de causale relatie tussen de interventie en het gevonden effect. *Ten derde* wordt per methode aangegeven op welk moment de controlegroep moet worden samengesteld. In sommige gevallen kan dit enkel aan de voorkant, voordat de interventie wordt uitgevoerd. Andere methoden kunnen ook gebruikt worden tijdens of na de uitvoer van de interventie. *Tot slot* wordt besproken in hoeverre de methode een ethisch dilemma met zich meebrengt. Een ethisch dilemma ontstaat wanneer de controlegroep wordt buitengesloten van een interventie, terwijl zij wel baat kunnen hebben bij deze interventie. Ook wordt per methode aangegeven aan welke randvoorwaarden moet worden voldaan om deze methode in te kunnen zetten. Zonder die essentiële randvoorwaarden is het niet mogelijk de methode in te zetten.

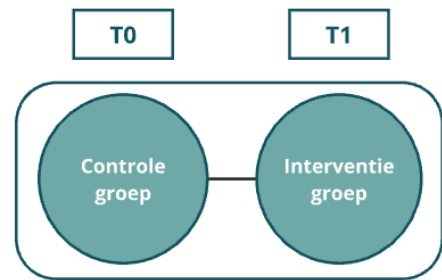
Aan uitvoerbaarheid van een methode en robuustheid van het bewijs wordt een score gekoppeld om deze elementen eenvoudig te vergelijken. De scores lopen van 1 tot 4. In het geval van uitvoerbaarheid betekent een score 1 dat deze methode eenvoudig uitvoerbaar is, score 4 betekent dat de methode lastig uitvoerbaar is. In het geval van robuustheid betekent een score 1 een lage mate van robuustheid, een score 4 betekent dan uiteraard een hoge mate van robuustheid. In onderstaande tabel is een samenvatting te zien van de vier facetten waarop de methodes met elkaar zijn vergeleken inclusief belangrijke randvoorwaarden voor inzet.



Methode	Uitvoerbaarheid	Robuustheid van bewijs	Moment van samenstellen	Ethisch dilemma	Randvoorwaarden voor inzet
Controlegroep = interventiegroep	● ○ ○ ○	● ○ ○ ○	Enkel vooraf	Geen	Groep waarbij de meting wordt uitgevoerd moet hetzelfde blijven tussen T0 & T1.
Vergelijking met een grotere groep	● ○ ○ ○	● ● ○ ○	Vooraf, tijdens of na afloop	Geen	Noodzakelijk toegang te hebben tot data van een grotere groep.
<i>Matching</i> op groepsniveau	● ● ○ ○	● ● ● ○	Enkel vooraf	Hoog	Er zijn twee groepen nodig die goed vergelijkbaar zijn met elkaar.
<i>Matching</i> op persoonsniveau	● ● ● ●	● ● ● ○	Vooraf, tijdens of na afloop	Laag	Toegang nodig tot de beoogde kenmerken voor <i>matching</i> tussen controle- en interventiegroep
Randomisatie	● ● ● ○	● ● ● ●	Enkel vooraf	Hoog	Noodzakelijk dat een groep niet geheel mee mag doen aan een interventie zodat een deel de controlegroep kan vormen.

## 4. CONTROLEGROEP = INTERVENTIEGROEP

De voor- en nameting is een veelgebruikte methode om onderzoek te doen naar het effect van een interventie. In het geval van een voor- en nameting wordt de voormeting (ook wel: T0) gebruikt als een controlegroep ten opzichte van de nameting (ook wel: T1). Een bepaalde score in de nameting wordt afgezet tegen de score in de voormeting. Zo kan bepaald worden of er sprake is van een toename of afname aan het einde van de interventie.



### Voorbeeld: Bijlesprogramma

Om aan te kunnen tonen of het bijlesprogramma werkt, wordt naar het gemiddelde van de cijfers gekeken vóór deelname aan de interventie en opnieuw na een bepaalde tijd. Als er een hoger gemiddelde wordt gezien, suggereert dit dat het bijlesprogramma een effect heeft gehad. Het is echter lastig om externe factoren uit te sluiten die mogelijk ook invloed hadden op het resultaat. Denk hieraan een ouder die meer betrokken is geraakt en de tijd neemt om de lesstof door te nemen en uit te leggen.

Deze methode heeft als belangrijk voordeel dat de interventiegroep zelf als controlegroep fungeert. Daarmee is de controlegroep meteen gevormd. Het nadeel is echter dat, omdat er geen aparte controlegroep is die de interventie niet krijgt, het niet mogelijk is om met zekerheid vast te stellen dat de waargenomen verandering daadwerkelijk door de interventie komt en niet door andere factoren<sup>xiii</sup>. Dit maakt het lastig om een causaal verband te bewijzen. Omdat er niet gecontroleerd kan worden voor externe invloeden (zoals natuurlijke ontwikkeling of externe gebeurtenissen) is de interne validiteit van deze methode beperkt. En omdat deze methode niet in staat is het effect van enkel de interventie te isoleren, zijn de resultaten lastiger te vertalen naar de grotere populatie. Dit beperkt de externe validiteit. Daarom wordt deze methode door experts afgeraden als het de wens is een causaal bewijs toe te wijzen aan de interventie<sup>xiv</sup>.

Tabel 2 – Controlegroep = interventiegroep

Element	Score	Toelichting
<b>Uitvoerbaarheid</b>	● ○ ○ ○	Deze methode is eenvoudig uitvoerbaar. Zo hoeft er geen aparte controlegroep gevonden of gevormd te worden. Dezelfde groep fungeert zowel als controlegroep (voor deelname) als interventiegroep (na deelname).
<b>Robuustheid van het bewijs</b>	● ○ ○ ○	De methode heeft beperkingen voor de interne en externe validiteit. Een verschil tussen de voor- en nameting kan namelijk ook goed verklaard worden door externe factoren die invloed hebben op de groep in de tussentijd tussen T0 en T1.
<b>Moment van samenstellen controlegroep</b>	Enkel vooraf	Vooraf. Het is essentieel aan de voorkant (voordat je de interventie uitvoert) een voormeting uit te voeren. De resultaten van deze voormeting worden namelijk vergeleken met de resultaten uit de nameting om uitspraken te kunnen doen over het effect.
<b>Ethisch dilemma</b>	Geen	In dit geval wordt er geen onderscheid gemaakt tussen controlegroep en interventiegroep. Daarmee wordt de interventie niet onthouden aan een controlegroep. Dat maakt dat binnen deze methode geen sprake is van een ethisch dilemma.

**Wanneer te gebruiken:**

- Wanneer het lastig is om toegang te krijgen tot niet-deelnemers voor een controlegroep.
- Wanneer een organisatie vooral zelf een indicatie wil krijgen van het effect dat een interventie heeft om hiervan te kunnen leren.

**Randvoorwaarde:**

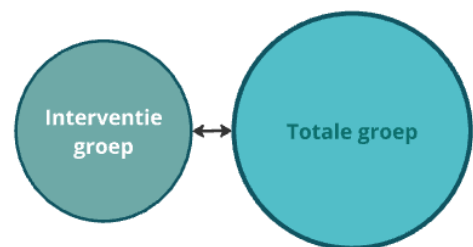
- Het is van belang dat de groep waarbij de meting wordt uitgevoerd hetzelfde blijft door de tijd heen. Als er gedurende de looptijd van het onderzoek veel mensen weggaan of juist nieuwe mensen bijkomen, zijn de resultaten vanuit de vergelijking tussen de T0 en T1 niet meer betrouwbaar.
- Het is belangrijk om binnen deze methode op persoon te matchen, zodat je kunt per persoon kunt zien of er sprake is van een ontwikkeling in de T1 ten opzichte van de T0.

## 5. VERGELIJKING MET GROTERE GROEP

Binnen de vergelijking met een grotere groep worden de resultaten van de interventiegroep afgezet tegen al bekende resultaten van een grotere totaalgroep.

Bij deze methode wordt de meting uitgevoerd binnen de groep die de interventie heeft ondergaan, idealiter wordt gewerkt met een voor- en nameting zodat inzicht wordt verkregen in of de interventiegroep aan de voorkant al hogere scores heeft in vergelijking met de grotere groep.

Vervolgens worden er bestaande landelijke of regionale gegevens gebruikt om de resultaten van de interventiegroep af te zetten tegen een controlegroep. De controlegroep wordt dus niet zelf samengesteld.

**Voorbeeld: Bijlesprogramma**

Het bijlesprogramma is specifiek ontwikkeld voor jongeren die voortijdig uit school dreigen te vallen in de regio Amsterdam. Het doel van het programma is om deze voortijdige uitval te voorkomen. De resultaten van het bijlesprogramma kunnen vervolgens vergeleken worden met de cijfers over voortijdig schoolverlaters op andere scholen in de regio Amsterdam. Wanneer de cijfers op de school waar het bijlesprogramma wordt gegeven lager zijn, is dat een indicatie dat het bijlesprogramma werkt om voortijdig schoolverlaten te verminderen.

Het voordeel van deze methode is dat de controlegroep niet beïnvloed kan zijn door de interventie. Wanneer de uitkomst bij de interventiegroep hoger is dan in de controlegroep is het wat zekerder dan de voorgaande methode dat de interventie hierin een rol speelt. Tegelijkertijd is niet met zekerheid te zeggen dat een hogere score bij de interventiegroep is toe te schrijven aan de interventie. Externe factoren kunnen binnen deze methode niet gecontroleerd worden, wat de interne validiteit beperkt. Wanneer niet zeker is of een effect is toe te schrijven aan een interventie, betekent dit ook dat er voorzichtig moet worden omgegaan met het generaliseren van de resultaten naar een bredere groep.

Tabel 3 – Vergelijking met een grotere groep

Element	Score	Toelichting
<b>Uitvoerbaarheid</b>	● ○ ○ ○	Deze methode is eenvoudig uitvoerbaar. Het is niet nodig om actief een controlegroep samen te stellen. Er wordt namelijk gebruikt van (secundaire) data die al beschikbaar is. Het is voldoende om enkel een meting bij de interventiegroep uit te voeren, de resultaten vervolgens te vergelijken met een grotere groep op basis van deze secundaire data.

<b>Robuustheid van het bewijs</b>	● ● ○ ○	De robuustheid van het bewijs dat deze methode oplevert is beperkt. Verschillen tussen de interventietroep en de grotere groep kunnen door meerdere externe factoren worden verklaard, waardoor met weinig zekerheid is aan te tonen of sprake is van een <i>extra</i> effect dat direct toewijsbaar is aan de interventie, ofwel de interne validiteit van deze methode is beperkt.
<b>Moment van samenstellen controlegroep</b>	Vooraf, tijdens of na afloop	Deze methode kan zowel tijdens de interventie, of nadat de interventie is afgerond worden ingezet, ervan uit gaande dat er (secundaire) data beschikbaar is.
<b>Ethisch dilemma</b>	Geen	Binnen deze methode worden geen groepen uitgesloten van de interventie, daarom is in dit geval geen sprake van een ethisch dilemma.

**Wanneer te gebruiken:**

- Wanneer het lastig is om toegang te krijgen tot niet-deelnemers voor een controlegroep.
- Wanneer er geen sterk robuust bewijs nodig is, bijvoorbeeld doordat de resultaten vooral gebruikt worden om een beeld te vormen van het effect van een interventie.

**Randvoorwaarde:**

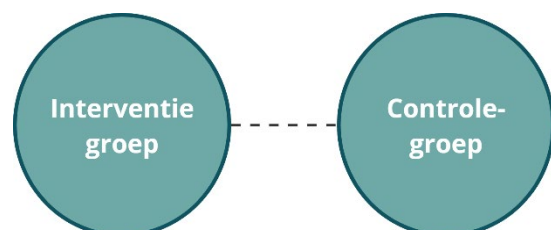
- Het is essentieel dat data die bij de interventiegroep wordt gemeten, ook verzameld wordt binnen de grotere groep. Wanneer centrale gegevens ontbreken, is vergelijking met een grotere groep niet mogelijk.
  - Het is aan te raden vooraf te verkennen welke gegevens beschikbaar zijn, zodat de eigen meting hier zo goed mogelijk op kan worden afgestemd.
- Er moet voor worden gezorgd dat de interventiegroep representatief is ten opzichte van de grotere groep die gebruikt wordt als controlegroep. Als deze twee groepen namelijk niet vergelijkbaar zijn, kunnen de resultaten ook niet vergeleken worden.

## 6. MATCHING OP GROEPSNIVEAU

De *matching* op groepsniveau<sup>5</sup> is een flexibele methode om een controlegroep samen te stellen. Voor de interventiegroep en controlegroep worden groepen gebruikt die al bestaan en die in veel opzichten op elkaar lijken.

De controlegroep binnen een *matching* op groepsniveau wordt gevormd door twee bestaande groepen te zoeken die zoveel mogelijk op elkaar lijken. Vervolgens ontvangt één van de twee groepen een interventie, de andere groep fungeert als controlegroep.

De controlegroep binnen deze methode is dus een bestaande, niet-gerandomiseerde groep die zoveel mogelijk lijkt op de interventiegroep. Er wordt geen nieuwe controlegroep geconstrueerd.



<sup>5</sup> In het Engels staat deze methode bekend als '*non-equivalent group design*'. In het Nederlands is de methode ook bekend onder de term 'quasi-experimenteel design' of 'verschilscoremethode', al zijn deze termen minder gangbaar.

### Voorbeeld: Bijlesprogramma

Een docent is benieuwd of een wekelijkse additionele bijles helpt om hogere cijfers te behalen. Er is niet direct de tijd om dit voor al haar klassen te doen. Daarom kiest ze in welke klas zij een extra bijlesmoment gaat inplannen om de jongeren te helpen. De ene klas krijgt dus een extra bijlesmoment en haar andere klassen krijgen dit niet. Vervolgens vergelijkt zij de cijfers van beide klassen met elkaar. De docent gebruikt bewust klassen van dezelfde school en hetzelfde jaar en de klassen hebben een vergelijkbaar gemiddeld startcijfer. Wanneer blijkt dat de klas die het extra bijlesmoment ontvangt een hoger gemiddelde heeft dan de andere klas, maakt dit het aannemelijk dat het extra bijlesmoment werkt om de cijfers van de jongeren te verhogen.

Binnen deze methode is er geen sprake van een willekeurige toewijzing aan controlegroep of interventiegroep. De groepen worden handmatig gekozen; vaak zijn dit al bestaande groepen die veel op elkaar lijken.

Binnen deze methode zijn niet alle karakteristieken van personen in de interventiegroep gekoppeld aan de controlegroep. De groepen zijn dus niet volledig gelijk aan elkaar. Dit betekent dat er een kans is dat er verschillen zijn die relevant zijn tussen beide groepen. Deze verschillen hebben invloed op het onderzoek, maar zijn niet bekend of zichtbaar. Dit beperkt de interne validiteit van het onderzoek. Een manier om dit risico deels te ondervangen is door gebruik te maken van een voormeting en nameting. De voormeting geeft voor een deel inzicht in eventuele verschillen die al aan de voorkant bestaan tussen interventie- en controlegroep.

Tabel 3 – *Matching* op groepsniveau

Element	Score	Toelichting
<b>Uitvoerbaarheid</b>	● ● ○ ○	Binnen deze methode wordt gebruik gemaakt van al bestaande groepen. Wanneer deze groepen voorhanden zijn (bijvoorbeeld klassen, of afdelingen binnen een bedrijf) is de methode redelijk eenvoudig toe te passen.
<b>Robuustheid van het bewijs</b>	● ● ● ○	De verschillen in uitkomsten tussen de interventiegroep en controlegroep kunnen verklaard worden door al bestaande, onderliggende en moeilijk observeerbare verschillen tussen beide groepen, die kans is hier wel kleiner dan bij de voorgaande methodes. Deze verschillen hebben invloed op de interne validiteit, en daardoor is de harde claim dat de interventie leidt tot het gewenste effect moeilijk tot niet te maken.
<b>Moment van samenstellen controlegroep</b>	Vooraf, tijdens of na afloop	Deze methode kan zowel met als zonder voormeting worden gebruikt. Zonder voormeting is toepassing mogelijk tijdens of na afloop van de interventie. Dit heeft wel gevolgen voor de robuustheid van het bewijs. Wanneer wél gebruik wordt gemaakt van een voormeting, dient deze methode voorafgaand aan de interventie te worden ingezet.
<b>Ethisch dilemma</b>	Hoog	Binnen deze methode wordt er een keuze gemaakt waarin de ene groep een interventie ontvangt en een andere groep deze interventie wordt onthouden. Dit brengt een ethisch dilemma met zich mee, zeker omdat hier bewust twee groepen worden gekozen. Het is belangrijk dat goed te kunnen verantwoorden. Ook is het mogelijk later een interventie alsnog aan te bieden aan de controlegroep waarmee het ethisch dilemma lager wordt.

**Wanneer te gebruiken**

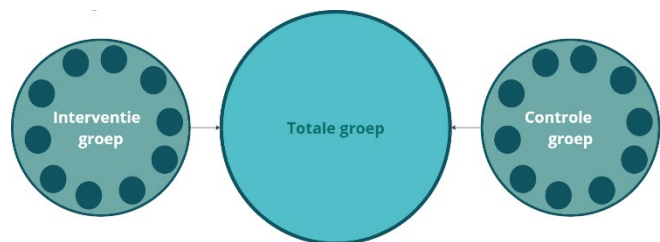
- In situaties waarin het samenstellen van een controlegroep door middel van willekeurige toewijzing niet haalbaar of wenselijk is, is de *matching* op groepsniveau vaak wel inzetbaar.

**Randvoorwaarde:**

- Het is nodig dat er toegang is tot twee groepen zijn die goed vergelijkbaar zijn met elkaar.
- Wanneer het de wens is de robuustheid zo sterk mogelijk te maken is het belangrijk te werken met een voormeting. In dat geval kan deze methode enkel voorafgaand aan de interventie worden ingezet.

## 7. MATCHING OP PERSOONSNIVEAU

*Matching* op persoonsniveau is een onderzoeksmethode waarbij deelnemers in de interventie- en controlegroep zo veel mogelijk op relevante kenmerken aan elkaar worden gekoppeld (denk aan kenmerken zoals leeftijd, geslacht of opleidingsniveau). Dit zorgt ervoor dat de groepen vergelijkbaar zijn, waardoor de invloed van versturende variabelen (zogenaamde *confounders*) wordt verminderd. Deze versturende variabelen kunnen het uiteindelijke effect van een interventie beïnvloeden.

**Voorbeeld: Bijlesprogramma**

Een docent wil onderzoeken of deelname aan een bijlesprogramma voor jongeren helpt bij het verbeteren van hun schoolprestaties. Als ze de prestaties van jongeren die deelnemen aan het programma vergelijkt met die van jongeren die dat niet doen, kunnen andere factoren, zoals thuissituatie of motivatie, het resultaat beïnvloeden. In plaats van de groepen direct te vergelijken, besluit ze *matching* toe te passen. Hierbij zoekt de docent jongeren die geen bijles hebben, maar die wel vergelijkbaar zijn met de deelnemers op bijvoorbeeld leeftijd, gezinssituatie en motivatie. Dit maakt de vergelijking eerlijker en helpt beter te begrijpen of het bijlesprogramma daadwerkelijk bijdraagt aan betere schoolprestaties.

*Matching* op persoonsniveau gebeurt door voor elke deelnemer aan een interventie iemand te zoeken die zoveel mogelijk op deze deelnemer lijkt. Om deze match te kunnen maken is het belangrijk dat relevante kenmerken bekend zijn. Het startpunt is dus de deelnemers aan de interventie. Vanuit deze interventiegroep wordt op basis van *matching* op persoonsniveau een controlegroep samengesteld die zoveel mogelijk overeenkomt met de kenmerken van de interventiegroep.

Tabel 4 – *Matching* op persoonsniveau

Element	Score	Toelichting
<b>Uitvoerbaarheid</b>	● ● ● ●	Deze methode is veeleisend om uit te voeren, al is de complexiteit deels afhankelijk van het aantal kenmerken dat wordt meegenomen in de <i>matching</i> van de controlegroep. Waar meer kenmerken de validiteit vergroten, verhoogt dit de benodigde middelen voor de uitvoering.
<b>Robuustheid van het bewijs</b>	● ● ● ○	Hoe robuust het bewijs is, hangt af hoeveel kenmerken worden meegenomen in de <i>matching</i> . Hoe meer variabelen zijn meegenomen in het maken van de matches, hoe robuuster het bewijs. Hiermee wordt immers de kans kleiner dat een verschil in uitkomst door een van deze variabelen wordt veroorzaakt, dit verhoogt de interne validiteit. Er blijven binnen <i>matching</i> op persoonsniveau onzichtbare verschillen tussen controle- en interventiegroep. Dit maakt dat een mogelijk geobserveerd effect in theorie ook verklaard kan worden door deze onzichtbare verschillen.
<b>Moment van samenstellen controlegroep</b>	Vooraf, tijdens of na afloop	Deze methode kan zowel voorafgaand, tijdens als na afronding van een interventie worden ingezet.
<b>Ethisch dilemma</b>	Laag	Bij deze methode worden geen groepen uitgesloten van de interventie om een controlegroep te kunnen vormen. De controlegroep wordt gevormd op basis van data. Dat betekent wel dat van de controlegroep persoonskenmerken gevraagd worden, terwijl zij hier geen profijt van hebben. Dat levert een lage mate van ethisch dilemma op.

#### Wanneer te gebruiken

- Deze methode kan zowel vooraf worden uitgedacht, of juist tijdens of achteraf worden ingezet om een controlegroep samen te stellen. Vooraf heeft de voorkeur, op die manier kan de interventiegroep nog bijgestuurd worden als het voor bepaalde mensen lastig is om matches te maken.
- Deze methode wordt vaak toegepast wanneer zich geen grote groep deelnemers aandient en er juist buiten de interventiegroep naar vergelijkbare personen gezocht moet worden. Een aandachtspunt hierbij is dat sommige persoonlijke kenmerken moeilijk te meten en te matchen zijn.
- Deze methode kan goed gebruikt worden om robuust bewijs te verzamelen, waarmee met een hoge mate van zekerheid kan worden vastgesteld of effecten veroorzaakt worden door de interventie. Wel is belangrijk te realiseren dat voor de uitvoerbaarheid van deze methode relatief veel middelen nodig zijn. Dat vraagt om een afweging in hoeverre de robuustheid van het bewijs deze hoge inzet van middelen verantwoordt.

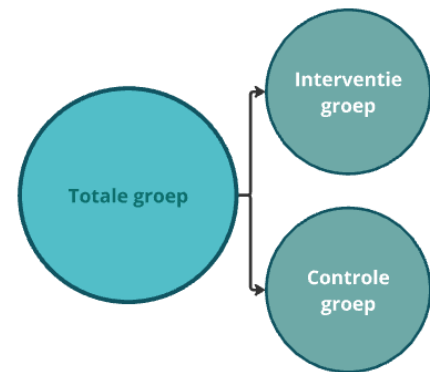
#### Randvoorwaarde:

- Om deze methode te kunnen gebruiken is het nodig om toegang te hebben tot de benodigde kenmerken, zoals opleidingsniveau, sociaaleconomische status of geboorteplaats op basis waarvan matches gemaakt kunnen worden tussen de interventiegroep en controlegroep.

## 8. RANDOMISATIE

In het geval van randomisatie worden deelnemers aan de interventiegroep en controlegroep willekeurig ingedeeld vanuit een grotere groep. De deelnemers hebben zelf geen enkele invloed op de groep waarin zij worden ingedeeld. Iedereen heeft dezelfde kans om geselecteerd te worden voor de interventiegroep of controlegroep. Randomiseren is een kernonderdeel van de *randomized controlled trial (RCT)*, een bekende en breed gebruikte methode.

Randomisatie zorgt voor een zeer hoge mate van vergelijkbaarheid tussen interventiegroep en controlegroep. Het enige verschil tussen beide groepen is in theorie immers dat deelnemers aan de controlegroep geen interventie ontvangen; deelnemers aan de interventiegroep wel.



### Voorbeeld: Bijlesprogramma

Een docent wil onderzoek of deelname aan een bijlesprogramma jongeren helpt om hun schoolprestaties te verbeteren. De docent wil voorkomen dat bepaalde zichtbare en onzichtbare factoren zoals thuissituatie, motivatie of concentratieproblemen het resultaat beïnvloedt. Daarom kiest ze ervoor om jongeren van de school willekeurig te selecteren op basis van hun leerling-nummer. De ene groep neemt deel aan het bijlesprogramma, de andere groep niet. Door het gemiddelde cijfer van beide groepen te vergelijken wordt inzicht verkregen of het bijlesprogramma daadwerkelijk bijdraagt aan betere schoolprestaties.

Randomisatie wordt gezien als een zeer betrouwbare manier om een controle- en interventiegroep samen te stellen. Omdat de controlegroep en interventiegroep vergelijkbaar zijn met elkaar is de kans klein dat andere variabelen de uitkomsten beïnvloeden<sup>xv</sup>. In het geval van randomisatie is de aanname dat factoren die eventueel ook invloed hebben op het uiteindelijke effect gelijk verdeeld zijn tussen controlegroep en interventiegroep. Als één van de factoren inderdaad invloed heeft op het effect, geldt dat voor beide groepen waardoor de uitkomsten voor het *extra* effect van de interventie nog steeds betrouwbaar zijn.

Randomisatie kan op verschillende wijzen worden uitgevoerd. Drie veelvoorkomende manieren zijn:

- Willekeurige indeling: elk individu uit de populatie krijgt een random nummer toegewezen. Vervolgens worden op basis van de random uitgedeelde nummers een interventie- en controlegroep gevormd.
- Gestratificeerde randomisatie: Hierbij wordt de totale populatie eerst verdeeld in groepen gebaseerd op specifieke kenmerken zoals geslacht of opleidingsniveau. Vanuit deze groepen wordt vervolgens willekeurig toegewezen aan controle- en interventiegroep.
- *Pipeline* randomisatie: Binnen deze methode ontvangt uiteindelijk iedereen de interventie, enkel het moment van aanvang verschilt. Dit wordt willekeurig bepaald. Sommigen nemen vanaf de start deel, anderen op een later moment. Degenen die later de interventie ontvangen fungeren in eerste instantie als controlegroep.



Tabel 5 – Randomisatie

Element	Score	Toelichting
<b>Uitvoerbaarheid</b>	● ● ● ○	De complexiteit van deze methode zit vooral in de werving en samenstelling van de controlegroep. Wanneer deelnemers weten dat ze mogelijk geen toegang krijgen tot de interventie, kan dit hun bereidheid tot deelname verminderen. Dit risico is met name aanwezig wanneer deelnemers expliciet geïnformeerd worden over hun toewijzing aan de controlegroep. In situaties waarin deelnemers zich niet bewust zijn van de groepsindeling, speelt dit probleem niet of minder. Toch vereist het zorgvuldig omgaan met verwachtingen en transparantie richting deelnemers, om zowel ethische als methodologische redenen.
<b>Robuustheid van het bewijs</b>	● ● ● ●	Deze methode levert sterk robuust bewijs op. De interne validiteit van deze methode is doorgaans hoog omdat eventuele gevonden verschillen tussen controlegroep en interventiegroep zeer waarschijnlijk toegeschreven kunnen worden aan de interventie.
<b>Moment van samenstellen controlegroep</b>	Vooraf	De methode van randomisatie kan enkel voorafgaand aan de interventie worden ingezet. Het is niet mogelijk om randomisatie in te zetten nadat de interventie al heeft plaatsgevonden.
<b>Ethisch dilemma</b>	Hoog	Binnen deze methode wordt de controlegroep willekeurig gekozen, deze controlegroep wordt (in eerste instantie) buitengesloten van de interventie. Dit brengt een ethisch dilemma met zich mee, omdat de interventie de controlegroep kan baten. Het is daarom belangrijk de keuze voor deze methode goed te verantwoorden.  Overigens bestaat ook een tegengeluid tegen dit ethische dilemma. Namelijk, dat het blijven inzetten van interventies waarvan de impact niet is aangetoond eveneens onethisch is <sup>xvi</sup> .

### Wanneer te gebruiken

- Deze methode om een controlegroep samen te stellen is enkel vooraf te gebruiken. Er moet van tevoren worden bepaald wie in de interventiegroep en wie in de controlegroep deelneemt.
- Deze methode levert robuust bewijs op. Dit is waardevol om harde uitspraken te kunnen doen dat een bepaalde interventie leidt tot gewenste effecten. Dit is bijvoorbeeld belangrijk wanneer de wens leeft een interventie op te nemen in een databank met effectieve interventies. Ook zal dit bewijs overtuigender zijn voor externe stakeholders, in vergelijking met andere methodes.

### Randvoorwaarde

- Een belangrijke voorwaarde voor randomisatie is dat er een groep is die niet geheel mee mag doen aan een interventie. Dat maakt het mogelijk om de totale groep van potentiële deelnemers willekeurig te verdelen in controlegroep en interventiegroep.

## 9. CONCLUSIE EN AANBEVELINGEN

Deze *whitepaper* heeft inzicht gegeven in de methodes om met een controlegroep aan de slag te gaan, en heeft getoond wanneer en hoe het inzetten van een controlegroep waardevol kan zijn om impact aan te tonen. Dit is gedaan aan de hand van vijf methoden om controlegroepen te vormen en te gebruiken die veel worden toegepast in de praktijk:

1. Controlegroep = interventiegroep
2. Vergelijking met een grotere groep
3. *Matching* op groepsniveau
4. *Matching* op persoonsniveau
5. Randomisatie

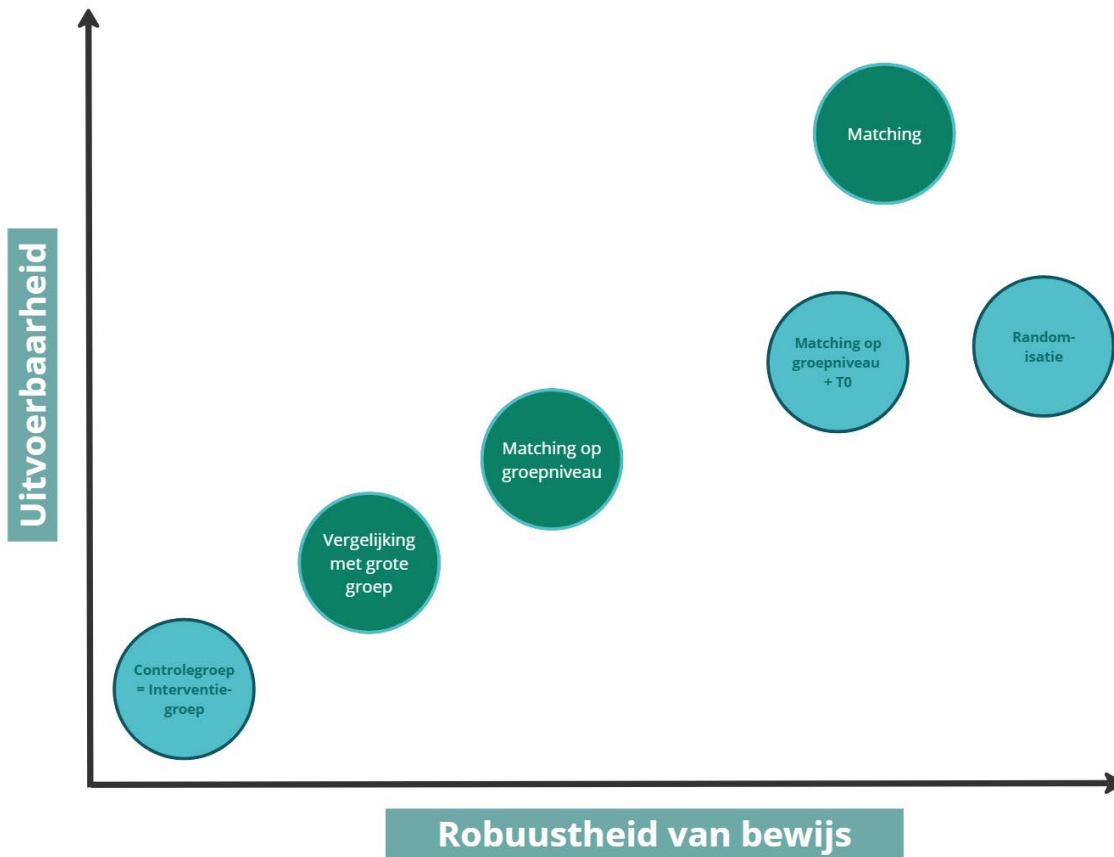
Iedere methode is vergeleken op basis van vier aspecten: de uitvoerbaarheid, de robuustheid van het bewijs, het moment van samenstellen van de controlegroep en een eventueel ethisch dilemma. Wat opvalt is dat de uitvoerbaarheid en de robuustheid van het bewijs aan elkaar gelinkt zijn. De methoden die eenvoudig uitvoerbaar zijn door hetzij geen separate controlegroep te hebben (*controlegroep = interventiegroep*) of de controlegroep niet zelf samen te stellen (*matching op groepsniveau*) laten tegelijk zien dat er een beperkte robuustheid is van het bewijs. Bij de andere drie methoden ligt de robuustheid van het bewijs hoger, maar dit vraagt ook meer in de uitvoerbaarheid. Tot slot speelt het ethische aspect bij *matching op groepsniveau* en randomisatie een belangrijke rol. Het verschil zit in de manier waarop de controlegroepen worden samengesteld. Bij randomisatie wordt dit willekeurig gedaan en is er dus geen bewuste keuze wie wel of niet bij de controlegroep of interventiegroep ingedeeld wordt. Bij *matching op groepsniveau* is dit andersom: hierin wordt wel een bewuste keuze gemaakt wie bij de controlegroep of interventiegroep wordt ingedeeld.

De vier facetten kunnen helpen om een passende keuze te maken voor de opzet van een controlegroep. Hier gaat het niet om welke de beste keuze is of welke keuzes geadviseerd worden, maar juist wat belangrijk is binnen de gegeven context. Mocht het ethisch aspect bijvoorbeeld als erg belangrijk worden geacht, dan is het logisch een methode te selecteren met een laag of geen ethisch dilemma. Terwijl wanneer het juist van belang is om een hoge robuustheid van bewijs te verkrijgen, het logisch is om voor een methode te gaan die dat realiseert. Welke methode het beste is, hangt af van de vraag van een organisatie en context.

In het onderstaande overzicht zijn de methoden ingedeeld op basis van 2 kenmerken:

- De sterkte van bewijs (horizontale as): Hoe verder een methode naar rechts staat, hoe sterker het bewijs dat daarmee wordt verzameld. Bij sterk(er) bewijs kunnen eventuele gevonden effecten directer worden toegeschreven aan de interventie.
- De uitvoerbaarheid van de methode (verticale as): Hoe hoger een methode in het schema staat, hoe lastiger de uitvoerbaarheid wordt omdat dit dan meer tijd, geld en expertise zal vragen om de methode uit te voeren.

Daarnaast is in het overzicht gebruikgemaakt van kleuren. De lichtblauwe bollen vertegenwoordigen methoden die alleen inzetbaar zijn wanneer er vooraf, vóór de start van de interventie al is bepaald wat de opzet is van de controlegroep. De donkergroene bollen staan voor methoden die eventueel ook tijdens of achteraf gebruikt kunnen worden.



Zoals te zien in de bovenstaand figuur zijn de methoden vrij goed verdeeld op een lineaire, diagonale lijn. Dat betekent dat er eigenlijk geen methodes zijn die zich linksboven bevinden. Dit zou ook een onpraktische categorie zijn van methodes aangezien deze methodes complex zijn om uit te voeren en dus veel middelen behoeven terwijl ze geen of zeer beperkt robuust bewijs opleveren. Ook rechtsonder is een gat zichtbaar. Er zijn dus ook geen methodes die een hoge robuustheid van bewijs opleveren en weinig complex zijn. Dit maakt dat er altijd op basis van de context van de meting een keuze gemaakt zal moeten worden die het meest geschikt is.

## 10. BRONVERMELDING

---

<sup>i</sup> Maas, K., Jong, N, de., Vermeulen, M., Kurpierz, S. "The application of baselines and control groups in PSI Monitoring & Evaluation". 2013.

<sup>ii</sup> Khandker, S. R., Koolwal, G. B., Samad, H. A. "Handbook on Impact Evaluation". 2010.

<sup>iii</sup> Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., Vermeersch, C. M. J. "Impact Evaluation in Practice". 2011.

<sup>iv</sup> Maas, K., Jong, N, de., Vermeulen, M., Kurpierz, S. "The application of baselines and control groups in PSI Monitoring & Evaluation". 2013.

<sup>v</sup> JMP Statistical discovery. (z.d.). *Correlation*. Statistics Knowledge Portal. Geraadpleegd op 18 juli 2025, van [Correlation | Introduction to Statistics | JMP](#)

<sup>vi</sup> Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. (2016). *Impact evaluation in practice*. World Bank Publications.

<sup>vii</sup> Khandker, S. R., Koolwal, G. B., Samad, H. A. "Handbook on Impact Evaluation". 2010.

<sup>viii</sup> Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. (2016). *Impact evaluation in practice*. World Bank Publications.

<sup>ix</sup> Khandker, S. R., Koolwal, G. B., Samad, H. A. "Handbook on Impact Evaluation". 2010.

<sup>x</sup> S. Morris, H Tödting-Schönhofer, and M Wiseman, "Design and Commissioning of Counterfactual Impact Evaluations," Publications Office of the European Union". 2013.

<sup>xi</sup> Morris, Tödting-Schönhofer, and Wiseman, "Design and Commissioning of Counterfactual Impact Evaluations."

<sup>xii</sup> Morris, Tödting-Schönhofer, and Wiseman; P.J. Ferraro, "Counterfactual Thinking and Impact Evaluation in Environmental Policy," Environmental Program and Policy Evaluation: Addressing Methodological Challenges. New Directions for Evaluation, 2009; K. Maas, M. Vermeulen, and S. Kurpierz, "The Application of Baselines and Control Groups in PSI Monitoring & Evaluation.," Erasmus University Rotterdam, 2013.

<sup>xiii</sup> T.R Knapp, "Why Is the One-Group Pretest-Posttest Design Still Used?," 2016.

<sup>xiv</sup> T.R Knapp, "Why Is the One-Group Pretest-Posttest Design Still Used?," 2016.

<sup>xv</sup> F. de Vocht et al., "Conceptualising Natural and Quasi Experiments in Public Health.," BMC medical research methodology, 2021.

<sup>xvi</sup> *Social Experiments to Fight Poverty / Esther Duflo*, 2010,  
[https://www.ted.com/talks/esther\\_duflo\\_social\\_experiments\\_to\\_fight\\_poverty](https://www.ted.com/talks/esther_duflo_social_experiments_to_fight_poverty).