

Experimenten Participatiewet – Weten wat werkt voor wie

Sanders*, M.W.J.L., Betko#, J. G., Blom-Stam+, K., Edzes§, A.J.E., Gramberg\$, P.J., Groot*, L., Kloosterman±, K., Kruijk*, M. De, Muffels+, R.J.A., Rosenkranz*, S., Rijnks[§], R.H., Spierings#, C.H.B.M., Swart\$, J.J.W. De, Venhorst[§], V.A., Verlaat*, T.

* Universiteit Utrecht, Utrecht School of Economics

Radboud Universiteit Nijmegen, Social Cultural Research

+ Universiteit Tilburg, TRANZO, Department of Sociology

§ Rijksuniversiteit Groningen, Economische Geografie

§ Hogeschool Saxion, Lectoraat Social Work

± Gemeente Groningen, Onderzoek, Informatie en Statistiek

1. Inleiding

Dit is de gezamenlijke oplegger van de kennisinstellingen bij de afzonderlijke evaluaties van de experimenten Participatiewet in de gemeenten Deventer, Groningen, Nijmegen, Tilburg, Wageningen en Utrecht. Deze kennisinstellingen zijn respectievelijk Saxion Hogeschool, Rijksuniversiteit Groningen, Radboud Universiteit, Universiteit Tilburg en Universiteit Utrecht. Daarbij is intensief samengewerkt in het door hen opgerichte Landelijk Overleg Experimenten Participatiewet (LOEP). Deze oplegger bevat de overkoepelende eerste conclusies die de onderzoekers trekken naar aanleiding van de experimenten.

Zes gemeenten hebben de afgelopen jaren (1 oktober 2017 – 31 december 2019) experimenten uitgevoerd in de bijstand onder het Tijdelijk Besluit Experimenten Participatiewet (Ministerie van Sociale Zaken en Werkgelegenheid, 2017). Het doel was om de uitvoering van de Participatiewet te verbeteren door minder controle en dwang en meer ruimte voor eigen regie, extra begeleiding en maatwerk. De gemeenten hebben de kennisinstellingen gevraagd de uitkomsten van deze experimenten wetenschappelijk te onderzoeken.¹ Daarbij zijn ze wetenschappelijk ondersteund door een Commissie van deskundigen van ZonMw. ZonMw heeft onafhankelijk een eigen proces evaluatie uitgevoerd. Daarnaast heeft het Ministerie van Sociale Zaken en Werkgelegenheid (SZW) het Centraal Planbureau (CPB) gevraagd een onafhankelijk onderzoek uit te voeren naar de resultaten op de uitkomstmaat uitstroom naar werk (De Boer et al., 2020). Om te kunnen vergelijken is in alle zes gemeenten door de onderzoekers grotendeels dezelfde vragenlijst afgenomen, zijn dezelfde uitkomstmaten gebruikt en is ook het analysemodel hetzelfde. Echter doordat de experimenten op gemeentelijk niveau zijn vormgegeven, zijn de verschillen tussen de gemeentes en de daar geïmplementeerde experimenten te groot om de data echt bij elkaar te kunnen voegen. Voor een volledige vergelijking en duiding van de zes experimenten verwijzen we naar de lokale eindverslagen en is verder vervolgonderzoek nodig. De experimenten leveren echter elk afzonderlijk maar vooral ook in samenhang relevante informatie op voor (toekomstig) beleid die de discussie over de toekomst van de sociale zekerheid in Nederland kan voeden. Daarom hebben we besloten deze gezamenlijke oplegger bij de rapporten te schrijven om een eerste inzicht in de bredere uitkomsten op een selectie van uitkomstmaten te geven. Na een korte samenvatting van de conclusies gaan we nader in op de verschillende experimenten en enkele belangrijke uitkomstmaten.

¹ De gemeenten vermelden in hun aanvragen in 2017 dat ze minstens zo groot belang hechten aan de uitkomsten op zelfredzaamheid, welbevinden en gezondheid als aan de in de AMvB genoemde "primaire uitkomstmaat", uitstroom naar betaald werk.

2. Samenvatting en conclusies

Alle experimenten hebben met elkaar gemeen dat de interventies (ten opzichte van het reguliere regime in de controlegroep) als “meer eigen regie, meer maatwerk en minder dwang” kunnen worden omschreven. De experimenten zijn op belangrijke punten anders vormgegeven (zie sectie 3). In deze oplegger vergelijken we de resultaten met de controlegroep (zie ook sectie 3).² Overeenkomstig de AMvB (Ministerie van SZW, 2017) zijn er in het experiment naast de reguliere dienstverlening (de vergelijkingsgroep) drie interventies (treatments):

1. Een ontheffing van de arbeids- en re-integratieplicht; *Zelf in Actie*.
2. Een intensievere en meer op maat gesneden begeleiding en ondersteuning; *Extra Hulp*.
3. Een ruimere bijverdienregeling voor inkomsten uit werk naast de uitkering; *Werken Loont*.

Doel van de Participatiewet is om mensen weer op eigen benen te laten staan en (snel) uit te laten stromen naar (betaald) werk of, als dat niet kan, op een andere manier te laten participeren. Kort samengevat laten de experimenten op de uitkomstmaat “uitstroom naar werk” een gemengd beeld zien. Daar waar in Utrecht veelal positieve effecten van de vrijstelling en extra hulp te zien zijn, zijn de effecten in Nijmegen in vergelijking met de controlegroep eerder negatief. Resultaten in Groningen, Deventer, Tilburg en Wageningen liggen daar ergens tussen en in alle gemeenten is de statistische significantie van de resultaten, positief of negatief, doorgaans gering.³ Maar ook als we concluderen dat op de meeste uitkomstmaten van uitstroom naar werk de effecten bescheiden zijn, is dit een belangrijke conclusie. Het betekent namelijk ook dat de huidige aanpak, met veel verplichtingen die met (dreiging van) strafkortingen worden afgedwongen, niet tot duidelijk betere resultaten leidt dan een aanpak die is gebaseerd op aandacht, vertrouwen, maatwerk en zelfredzaamheid van de deelnemer. Als dat het geval zou zijn, dan zouden in alle gemeenten de resultaten met andere aanpakken immers eenduidig negatief moeten uitpakken. En dat is niet het geval. We kunnen op basis van de experimenten niet één van de getestte interventies tot absolute winnaar uitroepen. Maar dat geldt dus evenzeer voor het huidige regime in de controle groepen.

Op de subjectieve uitkomstmaten, zoals zelfredzaamheid, welbevinden, participatie en gezondheid, verwachtten we op grond van de theorie eenduidiger positieve resultaten (Groot, Muffels en Verlaat, 2019). Immers, je zou verwachten dat bij de deelnemers meer vrijheid en autonomie, meer vertrouwen, meer bijverdiensten of gerichte hulp zou leiden tot betere rapportcijfers op deze aspecten. Maar ook hier is het beeld diffuus. In nagenoeg alle gemeenten overheerst het beeld dat er weinig significante positieve resultaten worden gevonden op de secundaire uitkomstmaten. Deels zijn deze verschillen te herleiden naar de verschillende opzet van de experimenten. Daarnaast zijn

² De resultaten in deze oplegger zijn gebaseerd op een lineair regressiemodel (zonder interacties) waarin gecorrigeerd wordt voor achtergrondkenmerken zoals leeftijd, geslacht, nationaliteit, opleiding, leefvorm, duur bijstand bij de start en looninkomen in de maanden voor start. Alleen in Groningen en alleen voor de uitkomstmaat uitstroom naar betaald werk wordt vergeleken met de gerandomiseerde referentiegroep. Vandaar dat we hierna spreken over de vergelijkingsgroep die in alle gemeenten behalve Groningen de controlegroep is.

³ In de gemeentelijke eindverslagen maar ook in het CPB rapport (De Boer et al., 2020) wordt opgemerkt dat de controlegroep in Nijmegen en in Tilburg al bij aanvang niet vergelijkbaar is met de overige interventiegroepen. Additionele analyses zoals gerapporteerd in de gemeentelijke rapportages, nuanceren de hier vermelde resultaten. Zo rapporteert Nijmegen kleinere of insignificante effecten na controle voor additionele variabelen en Tilburg en Wageningen rapporteren na gebruik van matching methoden en vergelijking met de referentiegroep ook bescheiden positieve effecten.

tal van factoren buiten het experiment van invloed op deze variabelen, zoals de gemeentelijke organisatie, de lokale omstandigheden op de arbeidsmarkt, en de wijze waarop het experiment is ingebed in de organisatie. De beperkte tijdsduur van de experimenten kan verklaren waarom de effecten op subjectieve uitkomstmaten als welbevinden en gezondheid (nog) niet voor alle gemeenten meetbare effecten laten zien.⁴ Een ander bekend fenomeen is dat de beleving van welbevinden, (mentale) gezondheid en sociaal vertrouwen niet snel veranderen (Uslaner, 2002) behalve door ingrijpende gebeurtenissen (zoals verlies van een naaste of een ernstige ziekte). Mensen passen zich tot slot ook relatief snel aan een nieuwe situatie aan waardoor bijvoorbeeld hun welbevinden en geluk weer snel naar de langetermijngemiddelden terugkeren (Headey et al., 2010). Daardoor zijn wellicht de effecten op met name de wat breder samengestelde uitkomstmaten als "welbevinden" en "gezondheid" bij de getoetste interventies klein. Op enkele variabelen, bijvoorbeeld zelfredzaamheid, werkzoekgedrag, sociaal vertrouwen, sociale participatie, zien we in sommige gemeenten wel positieve effecten.

Op basis van het geheel aan resultaten in onze experimenten kunnen we in ieder geval vaststellen dat de reguliere aanpak niet overal significant beter, en soms zelfs significant slechter werkt dan de interventies. Maar dat geldt dus evenzeer voor de geteste interventies. We concluderen daarom dat noch de geteste interventies noch het reguliere regime voor willekeurig samengestelde groepen bijstandsgerechtigden over alle gemeenten significant beter uitpakt.

Een mogelijke verklaring hiervoor is dat elke bijstandsgerechtigde een op maat gesneden aanpak vraagt. Met andere woorden, *one-size-does-not-fit-all*, terwijl de interventies voor alle deelnemers juist zoveel mogelijk gelijk waren, was in de controle groepen, waar het reguliere regime van kracht was, met het bestaande instrumentarium wel enige mate van maatwerk mogelijk. Dat is een mogelijke verklaring voor de bescheiden en soms negatieve effecten. De verscheidenheid van cliënten is eenvoudigweg te groot. Bij de interpretatie van deze resultaten is daarom van belang mee te wegen dat in de doelgroep een breed scala aan problemen met bijvoorbeeld armoede, sociale problematiek (o.a. in de gezinssituatie), mentale en fysieke gezondheid en gebrek aan opleiding (60% heeft geen startkwalificatie) samenkomen. De relatief bescheiden aantallen deelnemers maakt het niet goed mogelijk om voor al deze verschillen te controleren. Als zeer verschillende mensen met elkaar in een eenvormige interventiegroep (waar er extra is gelet op het aanbieden van dezelfde dienstverlening) komen, heeft wellicht een deel van de mensen baat bij die interventie, terwijl een ander deel juist minder positieve, of zelfs negatieve effecten laat zien. De effecten heffen elkaar dan op, met als resultaat kleine gemiddelde effecten met een hoge foutenmarge.

Daarnaast kunnen we niet uitsluiten dat er in sommige gemeenten experiment effecten zijn opgetreden. Dat wil zeggen dat bepaalde groepen beter of slechter presteren omdat deelnemers of consultants zich anders gedragen puur door deelname aan een experiment. Waar dit mogelijk van invloed is geweest wordt hiervan melding gemaakt (zie ook Box 2). Tot slot hebben de experimenten een relatief korte tijd mogen duren waardoor effecten nog niet uitgekristalliseerd zouden kunnen zijn. Voor de relatief korte termijn die het experiment heeft mogen duren is voor veel mensen, vooral

⁴ Zie ook de verschillende documenten van de begeleidingscommissie van ZonMW ter zake.

die met een grote afstand tot de arbeidsmarkt, (volledige) uitstroom naar betaald werk misschien wel erg ambitieus.

Samenvattend zijn de belangrijkste resultaten van deze experimenten bescheiden en vooral te vinden in wat ze *niet* laten zien. De interventies heffen verplichtingen op en geven extra aandacht en hulp. Wij vinden vervolgens geen bewijs dat het in stand laten van dwang en verplichtingen in de vergelijkingsgroepen leidt tot een hogere uitstroom naar werk. Het geven van méér ruimte aan gemeenten om te experimenteren met de aanpak en uitvoering van de bijstand zal dan ook niet leiden tot een significante verslechtering van die uitstroom, terwijl op langere termijn wel positieve effecten te verwachten zijn op het welbevinden van de mensen in de bijstand en dienstverleners. Door positieve aandacht en een intensievere en meer op maat gesneden begeleiding en ondersteuning zijn immers wel positieve resultaten gemeten op de subjectieve indicatoren, vooral op onderliggende variabelen als zelfredzaamheid, participatie en sociaal vertrouwen. De financiële ruimte om meer maatwerk en begeleiding te geven waar nodig kan deels gevonden worden door mensen die daar geen baat bij hebben juist vrij te stellen van dure, verplichte trajecten en de daarbij behorende controles. De ruimte om in deeltijd meer bij te verdienen naast de uitkering levert in een aantal gemeenten positieve en significante effecten op participatie en verdient serieus aandacht waarbij vanzelfsprekend rekening moet worden gehouden met de complexe effecten op toeslagen en subsidies. Interviews met deelnemers suggereren dat die complexiteit nog steeds een belangrijke belemmering voor het accepteren van werk in deeltijd vormt.

3. Overeenkomsten en verschillen tussen de zes experimenten

Voordat we naar de resultaten en uitkomstmaten gaan, is het van belang de verschillen tussen de experimenten goed te duiden. Niet elke gemeente heeft dezelfde interventiegroepen gebruikt en zelfs waar ze hetzelfde heten, verbergen de omschrijvingen soms nog behoorlijke verschillen in de implementatie. Zo verschillen de *Werken Loont*-regelingen en is vooral de *Extra Hulp*-interventie lokaal verschillend ingevuld. Toch is het mogelijk om twee 'panels' te creëren die ongeveer dezelfde interventiegroepen gebruikten. De gemeenten Groningen, Utrecht en Wageningen zitten in panel 1. Deze gemeenten hebben bijna uitsluitend pure interventies gebruikt. In panel 2 zitten de gemeenten Deventer, Nijmegen en Tilburg. Waarbij in de Deventer en Nijmegen juist uitsluitend en Tilburg vooral samengestelde interventies zijn getoetst. Een ander verschil in onderzoeksopzet betreft de werving en loting van de deelnemers. In Groningen zijn alle en in Deventer een gedeelte van de bijstandsgerechtigden willekeurig aan een interventie toegewezen en pas daarna uitgenodigd voor deelname, terwijl in de andere steden deelnemers pas na werving en aanmelding zijn toegewezen aan de interventies. Ten slotte is van belang mee te wegen dat zowel in de uitvoering als in de samenstelling van de doelgroep tussen de zes gemeenten verschillen bestaan die de uitkomsten kunnen hebben beïnvloed. Box 1 geeft kort weer waar men bij de duiding van resultaten in zogenaamde Randomized Control Trial (RCT)-experimenten extra op moet letten.

Box 1. Toelichting op de methode van onderzoek

De experimenten zijn opgezet als een RCT-experiment dat moet voldoen aan strikte onderzoekstechnische voorwaarden. Mensen worden daarbij (voor of na aanmelding) aselekt toegewezen aan de experimentgroepen. Bij de duiding van resultaten uit een dergelijk experiment zijn twee vragen belangrijk: (1) Kunnen de geobserveerde verschillen in uitkomsten tussen experimentgroepen uitsluitend worden toegeschreven aan een andere behandeling ofwel interventie (interne validiteit); (2) Gelden de geobserveerde effecten ook voor de doelpopulatie of andere populaties (externe validiteit)?

Hiervoor zijn drie aspecten belangrijk:

1. *Aselecte toewijzing*. Door randomisatie zijn de experimentgroepen in beginsel vergelijkbaar zowel op waargenomen (bijv. opleiding, duur bijstand) als niet-waargenomen kenmerken (bijv. gezondheid, motivatie) en startpositie op de uitkomstmaten. In de praktijk kunnen er toch toevallige verschillen zijn tussen de groepen die invloed kunnen hebben op de uitkomsten. Een uitsluitend causale interpretatie van gevonden verschillen in uitkomsten tussen groepen is dan niet meer mogelijk. Door statistische analysemethoden kan hiervoor deels worden gecorrigeerd.
2. *Geen post-randomisatie verschillen*. De wetenschap deel te nemen aan een experiment kan gedragseffecten hebben bij de deelnemers en consulenten (experimenteffecten). Daardoor kan de dienstverlening in de interventie- en of controlegroepen veranderen en kunnen effecten niet meer aan de interventie worden toegeschreven. Door vergelijking met een gerandomiseerde referentiegroep of door statistische analysemethoden (matching) kan voor experimenteffecten worden gecorrigeerd. Bij toewijzing ná aanmelding kan selectiviteit in de uitval ertoe leiden dat de groepen niet meer met elkaar te vergelijken zijn. Als voor uitvallers data beschikbaar is kan door statistische analysemethoden hiervoor deels gecorrigeerd worden.
3. *Selectiviteit van de steekproef*. Bij aanvang moet de gerandomiseerde controlegroep en interventiegroepen vergelijkbaar zijn met een bij voorkeur gerandomiseerde referentiegroep van niet-deelnemers. Is dat niet het geval, dan is er sprake van selectie bij deelnemers en/of niet-deelnemers en is de externe validiteit van gevonden effecten niet gewaarborgd. Bij toewijzing vóór aanmelding zijn de resultaten voor de experimentgroepen in hoge mate geldig voor de doelpopulatie van bijstandsgerechtigden, bij toewijzing na aanmelding is dit alleen het geval als de uitkomsten in de controle groep vergelijkbaar zijn met een gerandomiseerde referentiegroep.

Tabel 1 geeft in het kort de belangrijkste verschillen en overeenkomsten tussen de experimenten weer en in Box 2 vindt u per gemeente relevante bijzonderheden in de randomisatie en implementatie.

Tabel 1: Overeenkomsten en verschillen tussen de experimenten in de zes AMvB gemeenten

| | Deventer | Tilburg | Nijmegen | Utrecht | Groningen | Wageningen |
|--|--|--|--|--|--|---|
| Aantal deelnemers | 695 | 780 | 289 | 752 | 890 (8.724) ¹ | 410 |
| Looptijd experiment | 16-24 maanden | 16-24 maanden | 24 maanden | 16 maanden | 24 maanden | 24 maanden |
| Onderzoeksgroepen | (1) Zelf in Actie + Werken Loont (2) Extra Hulp+ Werken Loont (3) Zelf in Actie via App + Werken Loont | (1) Zelf in Actie + Werken Loont+ Werkbonus (2) Extra Hulp + Werken Loont + Werkbonus | (1) Zelf in Actie + Werken Loont (2) Extra Hulp+ Werken Loont | (1) Zelf in Actie (2) Extra Hulp (3) Werken Loont | (1) Zelf in Actie (2) Extra Hulp (3) Werken Loont (4) Keuze Treatment | (1) Zelf in Actie (2) Extra Hulp (3) Werken Loont |
| Ontheffing verplichtingen bij <i>Zelf in Actie</i> | Geen arbeids- en re-integratieplicht, wel informatieplicht. Maken van plan van aanpak. 1x per 6 maanden gesprek met vaste werkmakelaar en ondersteuning op eigen initiatief. | Geen arbeids- en re-integratieplicht, wel informatieplicht. Gebaseerd op 'Leren zelfredzaamheid' dus contact, zeker aan begin (2x per jr). | Geen arbeids- en re-integratieplicht, wel informatieplicht. Geen contact met deelnemers, tenzij zelf geïnitieerd. Alleen in combinatie met Werken Loont. | Geen arbeids- en re-integratieplicht, wel informatieplicht. Geen verplichte bijeenkomsten. Geen contact met deelnemers, tenzij zelf geïnitieerd. | Geen arbeids- en re-integratieplicht, wel informatieplicht. Geen verplichte bijeenkomsten. Geen contact met deelnemers, tenzij zelf geïnitieerd. | Geen arbeids- en re-integratieplicht, wel informatieplicht. Ondersteuningsmethode en 'Leren zelfredzaamheid'. |
| Extra dienstverlening bij <i>Extra Hulp</i> | Maken van plan van aanpak. Iedere 14 dagen contact met vaste werkmakelaar. | Gesprekken (6x per jr) met vaste coach, aanvullende instrumenten sociale domein en intensief contact met coach. | Voor dit experiment ontwikkelde begeleiding, met oa intensieve groepscoaching en veel autonomie deelnemers. Alleen in combinatie met Werken Loont. | Take-in gesprek met vaste werkcoach, aanvullende instrumenten en intensiever contact, ook op initiatief van de afdeling Werk en Inkomen. | Op maat dienstverlening door vijf extern aangetrokken ervaringsdeskundige begeleiders. | Treatment door vaste reguliere consultants, intensief contact en ondersteuning met instrumenten PW. |
| Vrijlating inkomsten bij <i>Werken Loont</i> | 25% tot €202. | 50% tot €202 i.p.v. 25% tot €202. | 50% tot €202 i.p.v. 25% tot €202. | 50% tot €202 i.p.v. 25% tot €202. | 50% tot €202 i.p.v. 25% tot €202. | 50% tot €202 i.p.v. 25% tot €202. |
| Werkbonus | | €2.392 (max. 2 keer uit te keren per half jr of jr) | | | | |

¹ Groningen heeft eerst de hele onderzoeksgroep (n=8.724) gerandomiseerd en daarna aan cliënten gevraagd of ze wilden deelnemen. Uiteindelijk hebben 890 deelnemers meegedaan.

Box 2: Methodische kanttekeningen per gemeente

Deventer

- *Randomisatie:* Vond plaats vóór aanmelding na opzij zetten van een willekeurig getrokken referentiegroep. Vanwege randomisatiemethode is vergelijking met de willekeurige referentiegroep zuiver. Doorgaans geen grote verschillen tussen experimentgroepen m.b.t. waarneembare achtergrondkenmerken.
- *Implementatie:* In de resultaten voor Extra Hulp is het effect van begeleiding door een groep vaste klantmanagers verweven.
- *Steekproef:* Niet iedereen mocht meedoen aan het experiment. Randomisatie vond plaats voor aanmelding. De deelnemers hebben een iets gunstiger arbeidsmarktpositie dan de doelgroep. Rond een kwart van de doelgroep deed mee aan het onderzoek.
- *Overig:* Door verschillende instroommomenten hebben deelnemers die vanaf het begin zijn ingestroomd vaker een vragenlijst ingevuld dan latere deelnemers.

Tilburg

- *Randomisatie:* Willekeurige groepsindeling vond plaats na aanmelding en is geslaagd.
- *Implementatie:* De controlegroep werd begeleid door consultants van de Afdeling Werk en Inkomen terwijl de treatmentgroepen werden begeleid door extern aangeworven coaches en projectleider. Zelf in Actie (met werkbonus) was 'het leren van zelfredzaamheid' van de deelnemers door daarvoor getrainde coaches en dus niet loslaten. De begeleiding was daardoor niet veel minder intensief dan bij de twee Extra Hulp groepen. De controlegroep lijkt daarentegen iets intensiever te zijn begeleid dan gangbaar. Er wordt melding gemaakt van mogelijke experiment effecten doordat controlegroep bij aanvang en direct daarna op uitstroom naar werk niet vergelijkbaar is met interventiegroepen. Effecten t.o.v controlegroep zijn daardoor niet toe te schrijven aan interventies.
- *Overig:* In het Tilburgse eindverslag is gebruik gemaakt van lokale BUS-data en zijn andere modellen gebruikt dan in dit verslag waardoor de resultaten iets kunnen afwijken. Er is middels propensity score matching gecontroleerd voor selectie in referentiegroep van alle niet-deelnemers in bijstand om uitstroom van interventiegroepen te kunnen vergelijken.

Nijmegen

- *Randomisatie:* Vond plaats ná aanmelding. Experimentgroepen zijn vergelijkbaar m.b.t. achtergrondkenmerken. In de maanden vóór start zijn er verschillen in de uitkomsten tussen de interventiegroepen en de controlegroep op het onderwerp 'deeltijdwerk', waarbij controlegroep kansrijker is. Bij aanvang zijn er (op basis van de vragenlijst) verschillen in oa (mentale) gezondheid, verslaving, en persoonlijke problemen, waarbij de controlegroep kansrijker is.
- *Implementatie:* In de resultaten voor Extra Hulp is het effect van begeleiding door een groepje vaste klantmanagers verweven.
- *Steekproef:* Niet iedereen mocht meedoen aan het experiment en randomisatie vond plaats na aanmelding. Als gevolg heeft de groep deelnemers een gunstigere arbeidsmarktpositie dan de doelgroep. Onderzoek wijst uit dat er binnen de potentiële deelnemers ook een verschil tussen mensen die zich opgaven en zij die dat niet deden, waarbij degenen die zich opgaven kansrijker waren.

Utrecht

- *Randomisatie:* Vond plaats ná aanmelding. Doorgaans geen grote verschillen tussen experimentgroepen m.b.t. waarneembare achtergrondkenmerken. In enkele maanden vóór start kleine verschillen in uitkomsten tussen de interventiegroepen en de controlegroep, waarbij controlegroep minder kansrijk.
- *Implementatie:* In resultaten voor Extra Hulp is het effect van begeleiding door een groep vaste klantmanagers verweven.
- *Steekproef:* Sommige groepen waren op voorhand uitgesloten van deelname; randomisatie vond plaats na aanmelding. Als gevolg heeft de groep deelnemers een gunstigere arbeidsmarktpositie dan de doelgroep.
- *Overig:* Uitkomsten uit de vragenlijsten voor Extra Hulp moeten i.v.m. een licht lagere respons voorzichtig worden geïnterpreteerd.

Groningen

- *Randomisatie:* Vond plaats vóór aanmelding na opzij zetten van een willekeurig getrokken referentiegroep. Vanwege randomisatiemethode is vergelijking met de willekeurige referentiegroep zuiver. Doorgaans geen grote verschillen tussen experimentgroepen m.b.t. waarneembare achtergrondkenmerken.
- *Implementatie:* Mogelijk gedragseffecten in de controlegroep waardoor vergelijking met deze groep niet zuiver is; waar mogelijk daarom vergelijking met willekeurige referentiegroep. Voor uitkomsten uit de vragenlijsten wel vergelijking met controlegroep omdat referentiegroep geen vragenlijsten heeft ingevuld.
- *Steekproef:* Door randomisatie van hele doelgroep hebben resultaten op basis van CBS data in hoge mate betrekking tot de doelgroep. Kortdurende bijstandsafhankelijken (< ½ jaar) horen niet tot de doelgroep.

Wageningen

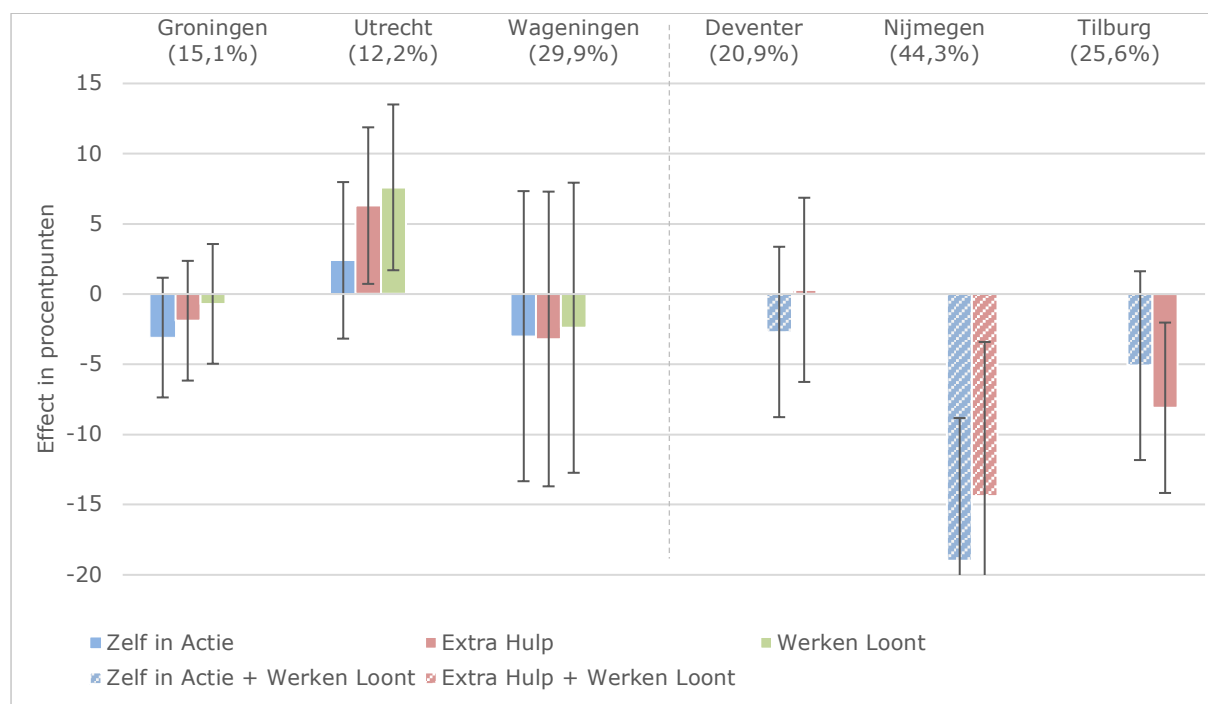
- *Randomisatie:* Vond plaats ná aanmelding. Experimentgroepen zijn vergelijkbaar m.b.t. achtergrondkenmerken. In sommige maanden voor start kleine verschillen in uitkomsten tussen de intensieve begeleidingsgroep en de controlegroep, waarbij controlegroep voor start kansrijker is voor kleine banen en meer dan 50% minimumloon.
- *Implementatie:* Voor alle groepen is een werkmethode ontwikkeld waardoor de begeleiding in de controlegroep intensiever was dan gangbaar. Er wordt melding gemaakt van mogelijke experimenteffecten, doordat de interventie in de controlegroep onder andere door de nieuwe methode na de start is veranderd.
- *Steekproef:* Niet iedereen mocht meedoen aan het experiment en randomisatie vond plaats na aanmelding. Rond de helft van de bijstandspopulatie deed mee aan het onderzoek.
- *Overig:* In het Wageningse eindverslag is gebruik gemaakt van lokale BUS-data en zijn andere modellen gebruikt dan in dit verslag waardoor de resultaten iets afwijken. Er is middels propensity score matching gecontroleerd voor selectie in referentiegroep van alle niet-deelnemers in bijstand om uitstroom te kunnen vergelijken.

4. Resultaten per uitkomstmaat

4.1. Uitstroom naar regulier betaald werk

Omdat de experimenten niet één-op-één met elkaar te vergelijken zijn, concentreren we ons in deze oplegger op tekens (positief of negatief) en significantie van gevonden effecten en geven we minder gewicht aan de geschatte grootte van de gemeten effecten.⁵ In alle onderzoeken is gekeken naar een breed scala aan uitstroommaten. In volgorde van een oplopend aantal uren is gekeken naar het vinden van werk voor meer dan 8, 12, 27 en 32 uur werk per week, een vast of tijdelijk contract en het verdienen van meer dan 50%, 70% en 100% van het wettelijk minimumloon (WML). Bij deze vergelijkingen is gebruik gemaakt van data van het CBS die over alle gemeenten vergelijkbaar zijn. We presenteren hier alleen een vergelijking van de resultaten voor de 27 en 8 uur werk (grote en kleine baantjes) en 70% wettelijk minimumloon (uitstroom criterium voor alleenstaanden). Voor meer detail en nuance verwijzen we naar de afzonderlijke rapportages van de deelnemende gemeenten en de analyse zoals door het CPB gemaakt voor SZW (De Boer et al., 2020).

Figuur 1: Uitstroom >8 uur per week



Toelichting: Treatmenteffecten t.o.v. de vergelijkingsgroep in procentpunten en 90% betrouwbaarheidsinterval. Uitkomsten voor de vergelijkingsgroep in procent staan boven de figuur. Zelf in Actie + Werken Loont in Tilburg is Zelf in Actie + Werken Loont + Werkbonus. Voor Groningen is de vergelijkingsgroep niet de controlegroep, maar een willekeurig samengestelde referentiegroep. Zie ook de rapportage voor Groningen.

Het gaat in alle gevallen om het effect in de laatste maand van het onderzoek en ten opzichte van de vergelijkingsgroep.⁶ In Figuur 1 wordt de uitstroom weergegeven per interventiegroep voor werk

⁵ Er zijn drie onderzoeksgroepen niet meegenomen in de outputfiguren. De onderzoeksgroepen ('Keuze voor treatment' (uit Groningen), 'Zelf in Actie via App' (uit Deventer) en 'Extra Hulp + Werken Loont + Werkbonus' (uit Tilburg) zijn maar in één gemeente uitgevoerd en daarom niet zinvol vergelijkbaar met andere gemeentes. U vindt de resultaten voor die specifieke groepen in de betreffende gemeentelijke eindrapportages.

⁶ Utrecht en Tilburg: maand 16; Nijmegen: maand 22; Overig: maand 24. Groningen: Vergelijking met willekeurig samengestelde referentiegroep i.p.v. controlegroep.

van meer dan 8 uur per week (kleine baantjes) ten opzichte van de vergelijkingsgroep (het zogenaamde treatmenteffect) met daarbij het betrouwbaarheidsinterval.⁷ De geschatte effecten zijn in de figuur met een balk aangegeven, terwijl de 90% betrouwbaarheidsintervallen met een donkere lijn zijn aangegeven. We merken hierbij op dat, door de soms kleine aantallen deelnemers, resultaten wel significant zijn bij minder strenge toetsing. Het gemeten verschil ligt waarschijnlijker in het positieve (negatieve) deel van de figuur naarmate er een groter deel van het betrouwbaarheidsinterval boven (beneden) de as ligt.

Bij kleine baantjes zien we voor Groningen en Wageningen negatieve, maar insignificante effecten voor de interventies waarbij deelnemers werden ontheven van verplichtingen (in blauw). In Utrecht zijn deze effecten positief maar niet significant. In combinatie met ruimere bijverdiensten zijn de gemeten effecten ook negatief in Nijmegen, Tilburg, Deventer, al zijn ze alleen in Nijmegen ook significant op 90% niveau. Voor elke gemeente is ook het gemiddelde uitstroompercentage van de controlegroep tussen haakjes weergegeven. Hierbij dient vermeld te worden dat door verschillen in bijvoorbeeld de randomisatie en de inhoud van de interventies de vergelijkbaarheid tussen gemeenten niet erg groot is.⁸

Bij de *Extra Hulp* groep, die in de verschillende gemeenten ook verschillend is geoperationaliseerd, worden in Groningen de negatieve effecten kleiner en blijven ze zowel in Groningen als Wageningen insignificant. Voor Utrecht zijn de effecten positief en nu ook significant bij 90%, terwijl in Deventer deze interventie in combinatie met ruimere bijverdienste een positief, maar insignificant effect heeft. In Tilburg en Nijmegen zijn de effecten op kleine baantjes in deze groep significant negatief op 90% niveau.⁹ In alle gemeenten lijken de ruimere bijverdiensten, soms gecombineerd met andere interventies, tot meer kleine baantjes te leiden dan de andere interventies. Voor Utrecht is dit effect het sterkste, maar voor geen van de gemeenten is het verschil significant.

Over de gehele linie genomen, kunnen we concluderen dat het ontheffen van sollicitatieplicht, het moeten accepteren van een aangeboden baan en andere plichten, in de meeste gemeenten leidt tot minder bijstandsgerechtigden met kleine baantjes. Dat kan betekenen dat bijstandsgerechtigden bij deze treatment minder snel een kleine baan accepteren, maar ook dat ze dergelijke baantjes minder gemakkelijk op eigen houtje vinden of er minder hard naar zoeken wanneer de verplichting wegvalt. Resultaten in bijvoorbeeld Utrecht laten ook zien dat mensen bij *Zelf in Actie* vaker een vast contract vinden, hetgeen zou suggereren dat de kwaliteit van de uitstroom stijgt. Het feit dat bij de groep met intensieve hulp de negatieve effecten kleiner en in Utrecht de positieve effecten sterker zijn, lijkt te ondersteunen dat mensen in de bijstand een andere manier van hulp en bejegening nodig

⁷ Het 90% betrouwbaarheidsinterval kan als volgt worden geïnterpreteerd: als het experiment 100 keer zou worden herhaald, dan zou het geschatte effect minstens 90 keer binnen dit interval liggen. Het werkelijke effect ligt dus redelijk zeker binnen dit interval.

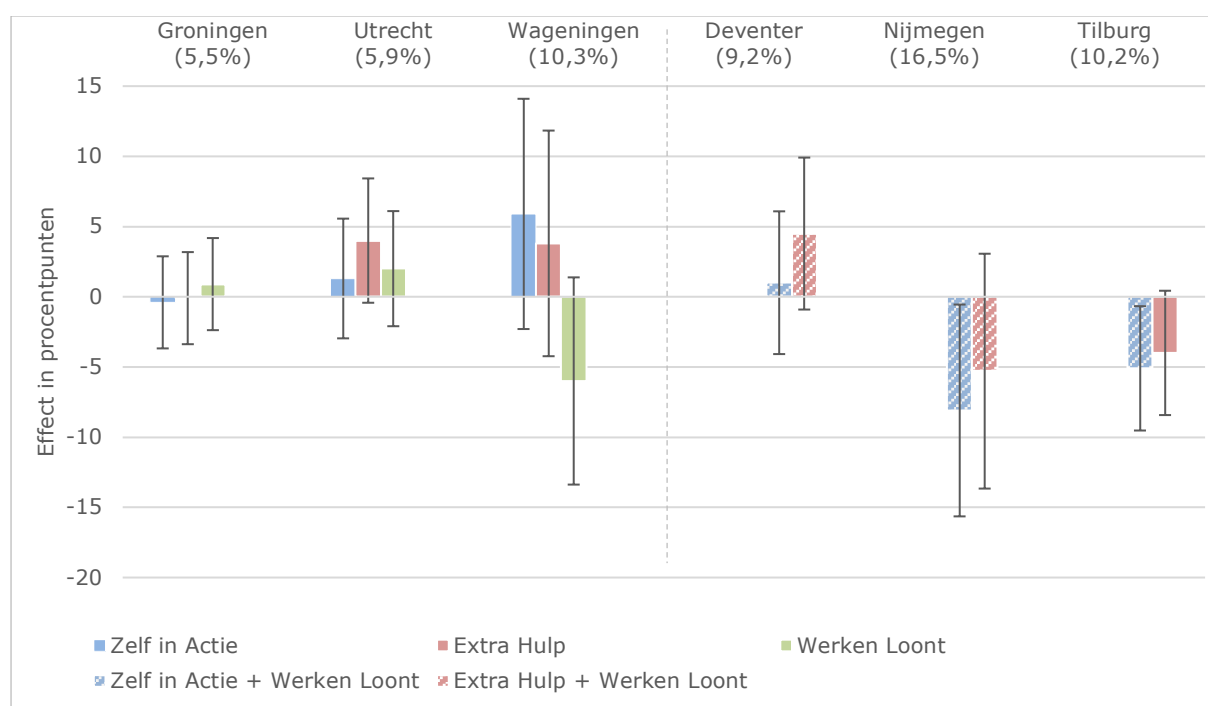
⁸ Een opvallend en belangrijk verschil tussen de gemeenten is dat in de controlegroepen, dus waar de reguliere dienstverlening gold, in Nijmegen in de laatste maand van het experiment 44% een baantje van meer dan 8 uur per week had, in Wageningen 30%, in Tilburg 26%, in Deventer 21%, in Groningen 15% en in Utrecht slechts 12%. De gemeten effecten zijn gemeten *ten opzichte van* die groep.

⁹ Hierbij moet de kanttekening worden gemaakt dat voor beide gemeenten volgens het CPB (De Boer et al., 2020) de controlegroep reeds bij aanvang verschilt en het negatieve effect dus niet kan worden toegeschreven aan de treatment. Ook in het eindverslag van Tilburg wordt hiervan melding gemaakt door verwijzing naar mogelijke experimenteffecten.

hebben bij het zoeken naar werk. Verder blijkt dat een ruimere bijverdienmogelijkheid bijstandsgerechtigden ook prikkel geeft om kleine baantjes te zoeken en accepteren.

Bij de grotere banen (>27 uur) zijn de percentages en effecten vanzelfsprekend kleiner in absolute zin (Figuur 2).¹⁰ In Tilburg valt op dat de *Zelf in Actie* groep significant minder grote banen vindt, waar dat voor de kleine banen nog niet zo sterk het geval was. In Wageningen zijn de effecten voor deze groep positief maar niet significant, net als in Deventer. In Utrecht en Groningen worden de positieve, respectievelijk negatieve effecten insignificant en in Nijmegen en Tilburg blijft het effect voor de *Zelf in Actie* groep significant negatief. De groepen *Extra Hulp* (met en zonder *Werken Loont*) zijn in Deventer, Utrecht en Wageningen positief insignificant, terwijl ze in Groningen en Nijmegen insignificant en in Tilburg significant negatief zijn.¹¹ Over de gehele linie worden de effecten minder negatief bij *Extra Hulp*. Ook hier is daarom de conclusie gerechtvaardigd dat de meeste bijstandsgerechtigden baat hebben bij *Extra Hulp* bij het vinden van werk. Nadere analyses uit Utrecht suggereren dat dit laatste effect vooral optreedt bij mensen met een grotere afstand tot de arbeidsmarkt.

Figuur 2: Uitstroom >27 uur per week

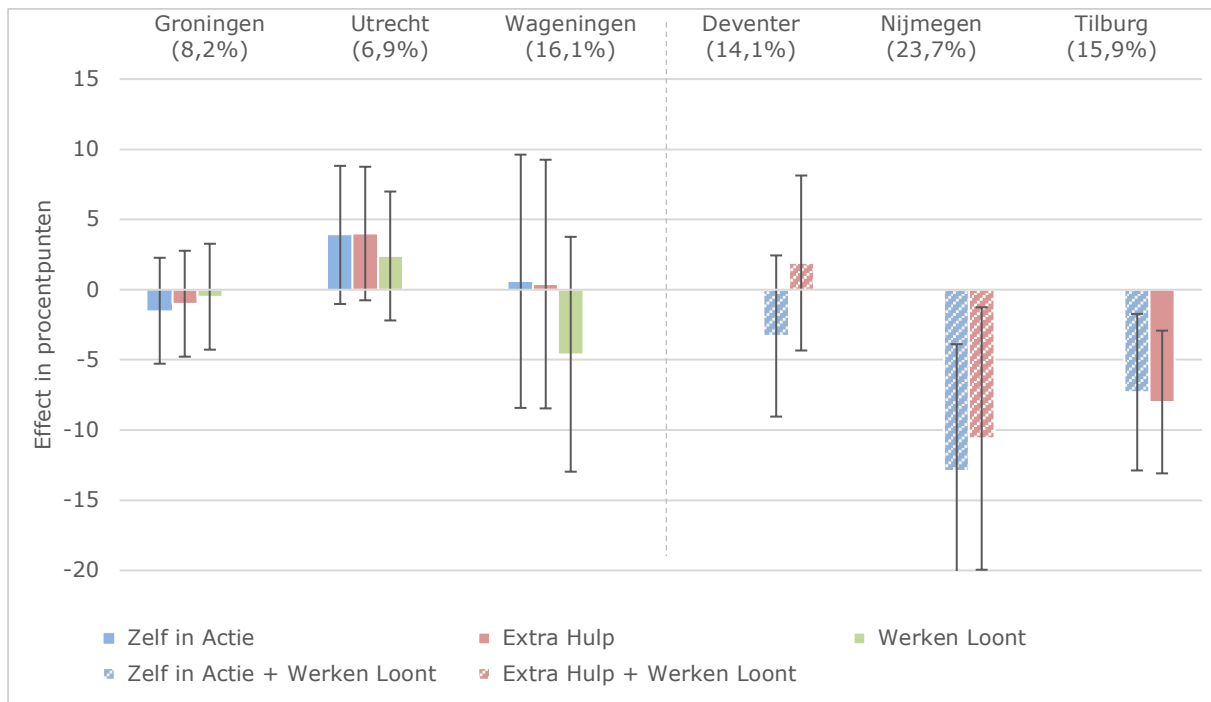


Toelichting: Treatmenteffecten t.o.v. de vergelijkingsgroep in procentpunten en 90% betrouwbaarheidsinterval. Uitkomsten voor de vergelijkingsgroep in procent staan boven de figuur. Zelf in Actie + Werken Loont in Tilburg is Zelf in Actie + Werken Loont + Werkbonus. Voor Groningen is de vergelijkingsgroep niet de controlegroep, maar een willekeurig samengestelde referentiegroep. Zie ook de rapportage voor Groningen.

¹⁰ Voor de controlegroep gelden percentages van 17% in Nijmegen, 10% in Tilburg en Wageningen, 9% in Deventer en 6% in Groningen en Utrecht. Per definitie zijn deze percentages lager dan voor baantjes van meer dan 8 uur, maar het patroon is hetzelfde.

¹¹ Ook hier geldt hetgeen reeds werd vermeld in voetnoot 14. Mogelijke selectieeffecten in de controlegroep kunnen voor Nijmegen en Tilburg dit resultaat vertekenen.

Figuur 3: Uitstroom naar >70% WML



Toelichting: Treatmenteffecten t.o.v. de vergelijkingsgroep in procentpunten en 90% betrouwbaarheidsinterval. Uitkomsten voor de vergelijkingsgroep in procenten staan boven de figuur. In Tilburg hebben beide combinatie treatments ook een Werkbonus. Voor Groningen is de vergelijkingsgroep niet de controlegroep, maar de willekeurig samengestelde referentiegroep. Zie ook de rapportage voor Groningen.

Als we kijken naar het verdienen van tenminste 70% van het wettelijk minimumloon (Figuur 3), dan worden de effecten, als verwacht, nog minder sterk en uitgesproken.¹² In Utrecht, Groningen, Wageningen en Deventer blijven de effecten veelal insignificant, terwijl in Nijmegen en Tilburg de negatieve effecten significant blijven, zowel voor de *Zelf in Actie* als voor de *Extra Hulp* groepen. Ook bij deze uitkomstmaat is het beeld dus wisselend en vooral door de resultaten in Tilburg en Nijmegen negatief. In Nijmegen geldt echter dat nadere analyse van de samenstelling van de controlegroep heeft uitgewezen dat daar, waarschijnlijk toevallig, toch vooral meer kansrijke deelnemers in de controlegroep terecht gekomen zijn (uitstroom naar 70% WML in de controle groep was in Nijmegen dan ook 24%, terwijl die landelijk en in de andere gemeenten tussen de 7-16% schommelt). Voor Tilburg geldt voor alle uitkomsten dat mogelijk sprake is geweest van een experimenteel effect waardoor de controlegroep anders is behandeld en zich wellicht ook anders is gaan gedragen. Daardoor kunnen de negatieve effecten niet worden toegeschreven aan de treatment (zie ook de Boer et al., 2020).

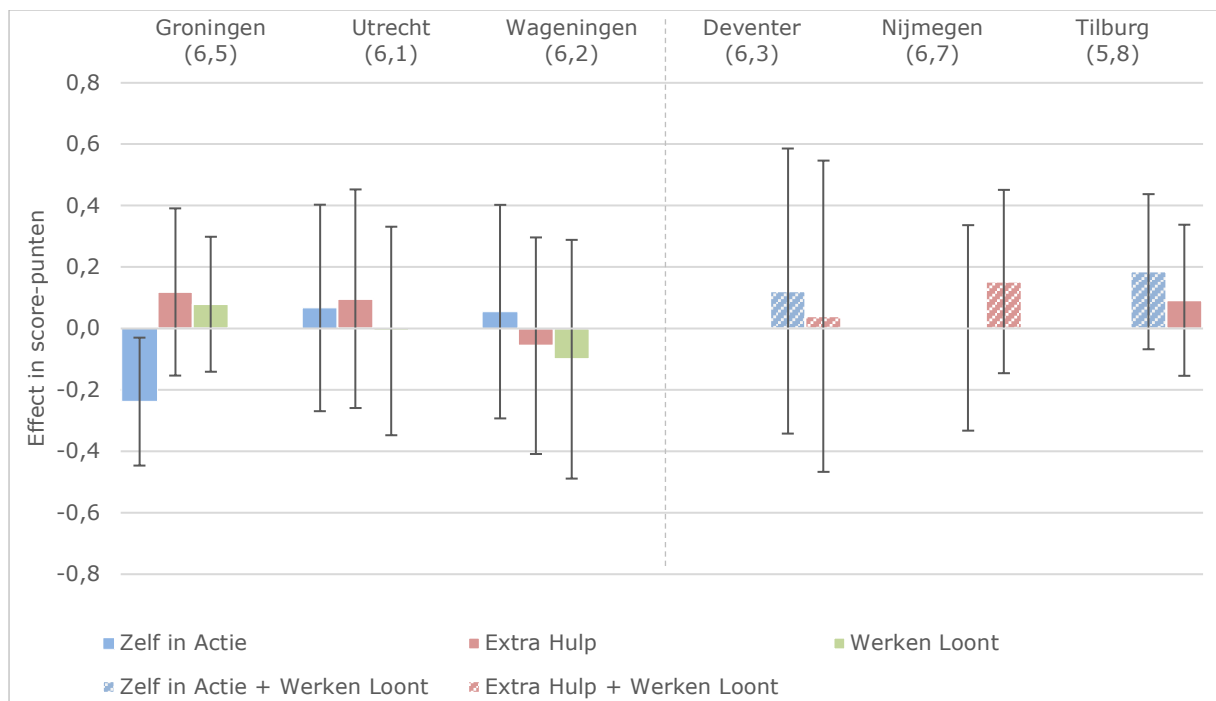
¹² Ook hier geldt dat er opvallende verschillen in de vergelijkingsgroepen zijn. In Utrecht heeft 7%, in Groningen 8%, in Deventer 14%, in Tilburg en Wageningen 16% en in Nijmegen 24% een inkomen boven de 70% van het WML in de laatste maand van het experiment. Uiteraard zijn deze percentages lager dan die voor baantjes van 8 uur of meer, maar opvallend genoeg hoger dan werk voor 27 uur of meer. Blijkbaar hoeft iemand dus geen 27 uur te werken om toch 70% WML te verdienen.

4.2. Welbevinden en vertrouwen in eigen kunnen

Voor het meten van subjectieve uitkomstmaten, zoals welbevinden, gezondheid en de ervaren kwaliteit van dienstverlening, is gebruik gemaakt van surveys. De eerste survey is afgenomen voor aanvang of kort na start van deelname in het experiment en diende als nulmeting. In de figuren hieronder ziet u de verschillen met de controlegroep op drie van de ongeveer 10 uitkomstmaten die zijn samengesteld op basis van de antwoorden uit de surveys in de laatste meting in het onderzoek.¹³ Door aan te sluiten bij de literatuur en items in de survey over alle gemeenten af te stemmen, zijn resultaten vergelijkbaar. Voor een verantwoording van de gebruikte vragen en uitgebreide resultaten verwijzen we u naar de onderliggende rapportages. We bespreken hier een selectie van de resultaten op "welbevinden" en de variabelen "sociaal vertrouwen" en "werkzoekgedrag".

Bij de scores op "welbevinden" in Figuur 4 valt op dat geen van de gemeenten significante positieve of negatieve effecten laten zien, met uitzondering van de groep *Zelf in Actie* in Groningen.¹⁴ De gemiddelde scores voor de controlegroepen variëren hier licht tussen de 5,8 (Tilburg) en 6,7 (Nijmegen).

Figuur 3: Welbevinden (score 0-10) laatste maand



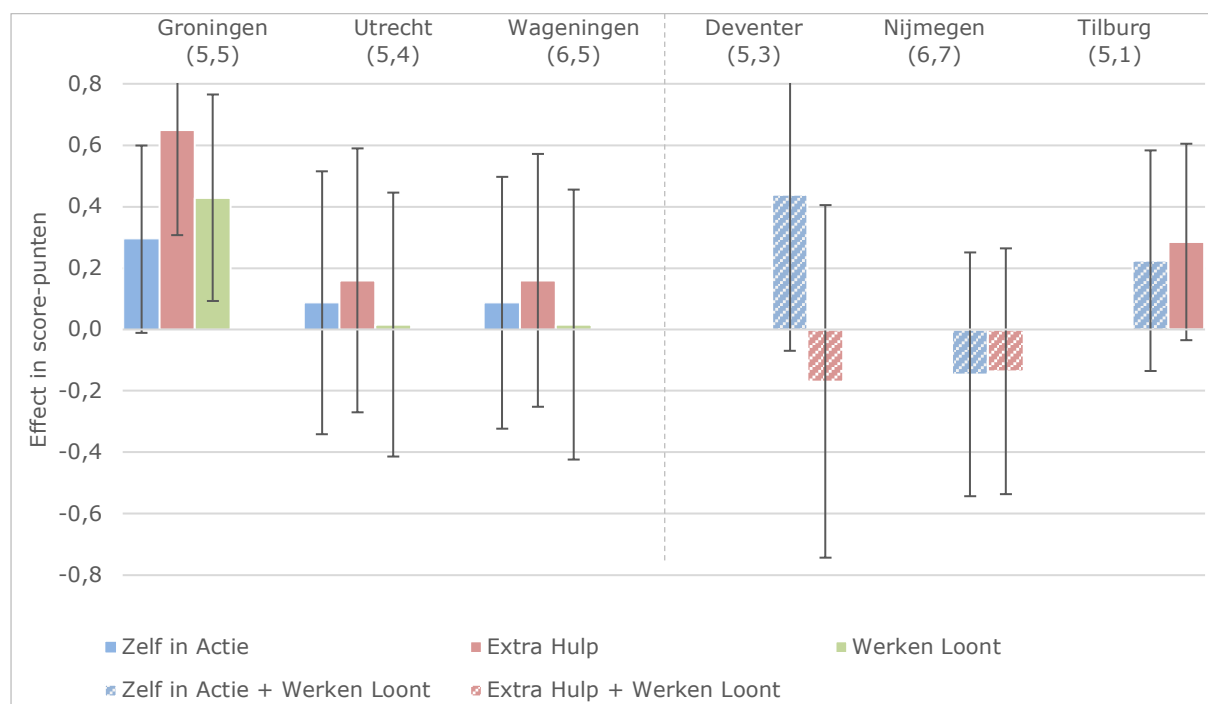
Toelichting: Treatmenteffecten t.o.v. de vergelijkingsgroep in procentpunten en 90% betrouwbaarheidsinterval. Uitkomsten voor de vergelijkingsgroep in procent staan boven de figuur. Zelf in Actie + Werken Loont in Tilburg is Zelf in Actie + Werken Loont + Werkbonus. Voor Groningen is de groep waarmee vergeleken wordt deze keer de controlegroep. Zie ook de rapportage voor Groningen.

¹³ In deze analyses is in alle gemeenten gecontroleerd voor het antwoord in de nulmeting. In Tilburg en Wageningen zijn de uitkomsten gebaseerd op het gemiddelde voor de twee vervolgmetingen.

¹⁴ Er zijn in individuele gemeenten wel, ook op deze index, licht positieve en soms significante effecten gemeten in de tussenmeting. De reden dat deze effecten in de laatste maand weer verdwijnen en/of minder uitgesproken zijn, kan deels liggen aan het vooruitzicht dat het experiment en daarmee de treatments zullen eindigen of het effect dat uitkomstmaten zoals welbevinden zelfs op redelijk korte termijn een terugkeer naar het gemiddelde vertonen.

Dit mag niet verbazen, aangezien we uit de literatuur ook weten dat de verschillen tussen mensen op deze indicator redelijk groot zijn, maar ook vrij stabiel over de tijd en dat gemiddelden over grote groepen dicht bij elkaar liggen (Kunzman et al., 2000; Helliwell and Huang, 2014; Headey e.a. 2010). Daarnaast is deze brede welzijnsindicator opgebouwd uit verschillende onderliggende variabelen uit de surveys (tevredenheid met leven, betekenis van leven), die eventuele verschillen tussen groepen mogelijk enigszins "uitmiddelen".

Figuur 5: Sociaal vertrouwen (score 0-10) laatste maand



Toelichting: Treatmenteffecten t.o.v. de vergelijkingsgroep in procentpunten en 90% betrouwbaarheidsinterval. Uitkomsten voor de vergelijkingsgroep in procent staan boven de figuur. Zelf in Actie + Werken Loont in Tilburg is Zelf in Actie + Werken Loont + Werkbonus. Voor Groningen is de groep waarmee vergeleken wordt deze keer de controlegroep. Zie ook de rapportage voor Groningen.

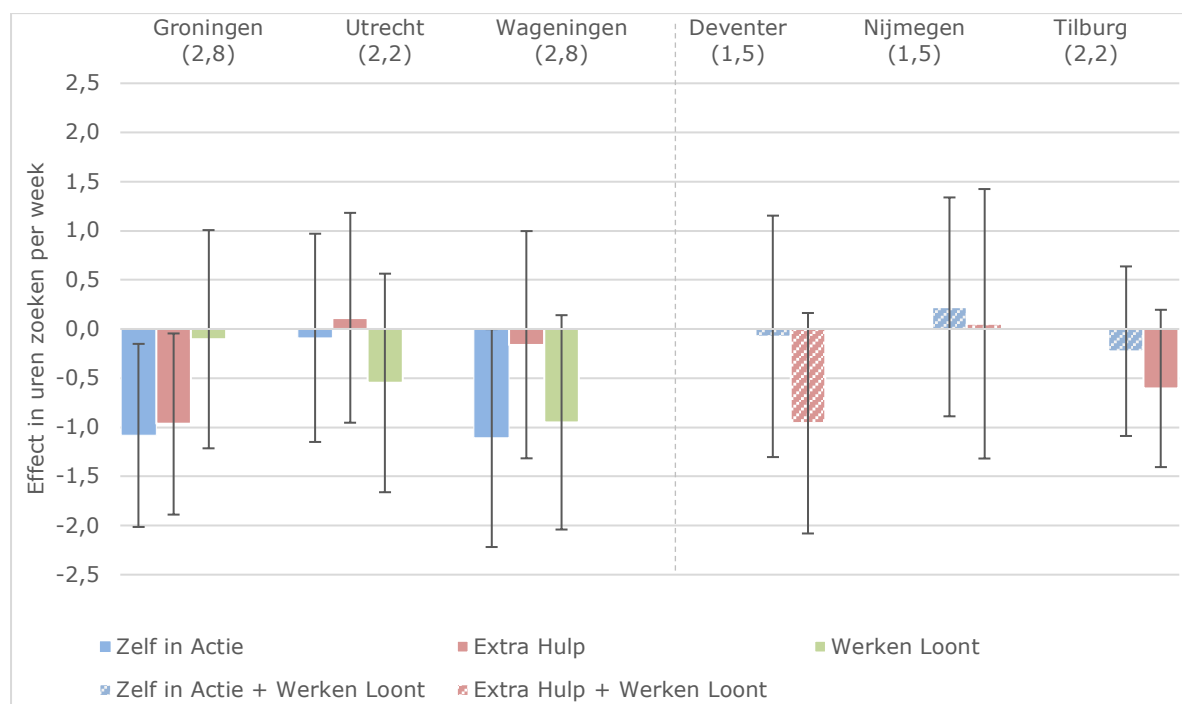
Als we kijken naar het "sociale vertrouwen" in Figuur 5 dan variëren de scores tussen de gemeenten van 5.1 in Tilburg tot 6.7 in Nijmegen in de controlegroep en zijn de effecten van de verschillende interventies klein en veelal insignificant. Alleen in Groningen registreren we een significant positief effect op deze variabele voor de *Extra Hulp* groep.

We vermelden hier ook de effecten op het werkzoekgedrag van de deelnemers. In de surveys is gevraagd naar het aantal uren per week dat actief is gezocht in de afgelopen vier weken. In de controlegroep varieert deze variabele van 1,5 in Deventer en Nijmegen tot 2,8 in Groningen en Wageningen.¹⁵ Je zou verwachten dat ontheffing van sollicitatieverplichtingen zou leiden tot minder

¹⁵ De variabele meet het gemiddelde aantal uren dat is gezocht naar betaald werk per week in de laatste maand.

zoeken, terwijl een hogere beloning voor werk zou leiden tot een hogere zoekintensiteit. De gemeten effecten staan afgebeeld in Figuur 6.¹⁶

Figuur 6: Werk gezocht (in uur/week) in de laatste maand



Toelichting: Treatmenteffecten t.o.v. de vergelijkgroep in procentpunten en 90% betrouwbaarheidsinterval. Uitkomsten voor de vergelijkgroep in procent staan boven de figuur. Zelf in Actie + Werken Loont in Tilburg is Zelf in Actie + Werken Loont + Werkbonus. Voor Groningen is de groep waarmee vergeleken wordt deze keer de controlegroep. Zie ook de rapportage voor Groningen.

We zien dat bij *Zelf in Actie* en *Extra Hulp* in Groningen de zoekintensiteit significant afneemt, terwijl dat in Deventer, Tilburg en Utrecht insignificant is. De grote verschillen in het effect van *Extra Hulp* op deze variabele (zie bijvoorbeeld Groningen en Wageningen) is wellicht een aanwijzing dat de invulling van de *Extra Hulp* interventie tussen de gemeenten verschilde in de nadruk op het zoeken naar betaald werk. In Nijmegen zijn de tekens opvallend genoeg positief, zij het insignificant. Belangrijk is hier ook te vermelden dat de effecten op het werkzoekgedrag zich zeker niet één op één vertalen naar effecten op werk vind- en uitstroomkansen. Waar bijvoorbeeld in Nijmegen de effecten op die maatstaven consistent negatief en vaak significant zijn, is het effect op zoekintensiteit eerder positief, terwijl dat beeld voor Utrecht, met uitzondering van *Extra Hulp*, eerder omgekeerd is. Dit onderstreept nog eens dat verplichte activiteiten om naar werk te zoeken (bv sollicitatieplicht) niet tot effectief zoekgedrag hoeft te leiden, terwijl een goede begeleiding het werkzoekproces juist efficiënter kan maken.

¹⁶ Deze maat vertoont wel een vertekening. De mensen die uitstromen vanwege werk, zoeken veelal niet meer en hebben bovendien een lagere respons in de latere metingen. Wat het netto effect van deze tegengestelde vertekeningen is, is niet te bepalen.

5. Discussie en Beleidsaanbevelingen

Er zijn wat ons betreft, zowel uit het proces als uit de resultaten, verschillende lessen te trekken. De eerste en wellicht belangrijkste les uit het proces is dat, wanneer er behoefte bestaat aan onderling goed vergelijkbare data, het van groot belang is om dit bij aanvang van het onderzoek vast te stellen. Daarbij horen dan ook voldoende middelen en ruimte voor een goede afstemming. De coördinatie tussen vijf onderzoeksgroepen die rapporteren aan zes projectgroepen in zes gemeenten, vergt veel van de betrokkenen. In deze zes lokale experimenten is dit prettig en productief verlopen, dankzij de inzet van velen en de wil om de experimenten goed uit te voeren. Maar voor de toekomst raden wij aan dit op voorhand en in de aanbesteding van het onderzoek goed te beleggen en ook te begroten. Hierbij is onze aanbeveling om vooral voor een uniform design wat betreft randomisatie, basisinterventies en implementatie te zorgen.

De tweede les heeft te maken met het karakter van een veldexperiment ofwel Randomized Controlled Trial. Bij het implementeren van een veldexperiment hebben we als onderzoekers te maken een complexe realiteit waarbinnen onze invloedssfeer beperkt is en we toch dienen te navigeren om het onderzoek gedegen te kunnen doen. Het bijzondere aan een RCT is dat dit een belangrijke methode is waarmee causaliteit van effecten met meer zekerheid kan worden aangetoond. Daaraan zijn dan wel heldere voorwaarden verbonden voor wat betreft de onderzoeksopzet. Daar schuurt het soms als die voorwaarden raken aan domeinen waarover wij als onderzoekers geen zeggenschap hebben. Daardoor kan het voorkomen dat de onderzoeksopzet ineens niet meer leidend is, maar andere factoren een rol spelen. Bijvoorbeeld als voor een willekeurige groep mensen een interventie waarvan de klantmanager overtuigd is dat deze goed of juist slecht zal werken, toch moet worden onthouden of juist geïmplementeerd. Of om niet tussentijds bij te sturen in een interventie, terwijl deze niet goed lijkt te werken. In dit specifieke geval hebben daarnaast ook de beperkingen die vanuit de landelijke (AMvB) en gemeentelijke (bijvoorbeeld samengestelde treatments) politiek zijn opgelegd, de bruikbaarheid van het onderzoek beperkt. Om voldoende "power" in het onderzoek te garanderen, bestaat er een afruil tussen de grootte van de interventies, de grootte van de groepen deelnemers en het aantal interventiegroepen. De politieke noodzaak om de interventies te beperken in grootte (bv. een marginaal ruimere bijverdienste), maar uit te breiden in aantal (bv. de verplichting om ook een intensiveringsgroep mee te nemen), loopt dan met de vrijwillige deelname al snel tegen de grenzen van de power op. We willen alle betrokkenen die op een moment de (politieke) verleiding of druk konden weerstaan om eisen of beperkingen op te leggen bij de totstandkoming van de experimenten bedanken en complimenteren voor hun (politieke) moed. We trekken hier ook zelf de les uit dat we als onderzoekers deze grenzen (nog) scherp(er) moeten bewaken en streng(er) moeten toezien op het zuiver en eenduidig implementeren van de interventies.

Ten slotte lijkt het ons raadzaam om in vervolgonderzoek de onderzoeksvraag te verleggen van "weten wat werkt" naar "weten wat werkt voor wie". De hier gekozen RCT-opzet is zeer geschikt voor onderzoek naar de effecten van interventies op de gemiddelde uitkomsten voor een representatieve groep deelnemers. Daarbij is van groot belang dat er voldoende deelnemers in de verschillende interventiegroepen deelnemen, zodat binnen die groepen nog onderscheid kan worden

gemaakt naar achtergrond- en andere kenmerken. Wat voor de een werkt, werkt voor de ander niet. Eerste voorzichtige analyses met subgroepen geven wel aan dat bijvoorbeeld in Utrecht lager opgeleiden *meer* baat hebben bij de *Zelf in Actie* én de *Extra Hulp* interventies en bijstandsgerechtigden met een grote afstand tot de arbeidsmarkt *alleen* bij *Extra Hulp*. Door te veel experimentgroepen in een experiment op te nemen, verwatert echter de power en kunnen dergelijke interessante verbijzonderingen niet meer statistisch betrouwbaar worden vastgesteld. Voor meer inzicht zou een vervolgonderzoek daarom eerst de heterogeniteit tussen de bijstandsgerechtigden goed in kaart moeten brengen, zodat keuzes kunnen worden gemaakt voor welke doelgroepen een specifieke interventie wordt getest. Gezien de korte tijdsspanne van het onderzoek, in sommige gemeenten gebaseerd op resultaten over 16 maanden, bestaat er behoefte aan vervolgonderzoek over een langere periode en aan duiding van de verschillen in resultaten tussen de gemeenten. De resultaten hier gepresenteerd zijn de eerste bevindingen omdat de uitkomsten nog nader onderzoek vragen. Wij roepen op het onderzoek in de toekomst in die richting verder te ontwikkelen.

Als we dan komen tot wat onze resultaten concreet kunnen betekenen voor het beleid, dan zijn we gepast bescheiden. De 'harde' resultaten wijzen eenvoudigweg niet sterk de ene of de andere kant op. Onze experimenten leren niet dat de ene of de andere interventie nu nationaal en over alle bijstandsgerechtigden moet worden "uitgerold". Dat is enerzijds een gevolg van kleine en zeer gevarieerde interventies over een korte periode, maar geeft op basis van deze eerste bevindingen aan, dat de veronderstelde positieve effecten van de theoretische interventies bij toetsing tegenvallen. Of dat een 'theorie' of 'uitvoering' probleem is moet nader worden onderzocht. Het kan ook zijn dat de interventies een andere werkmethode van de dienstverleners vragen en dat de in jaren opgebouwde expertise in de 'strenge' verplichtende aanpak niet van de ene op de andere dag goed kan worden geïmplementeerd in de uitvoeringspraktijk. Die expertise moet wellicht eerst worden opgebouwd. Als we onze resultaten leggen naast die van het SCP, dat in 2019 een uitgebreide evaluatie van de Participatiewet publiceerde, dan bevestigen deze het beeld dat het in de Participatiewet tijd is voor een (terugkeer naar een) menselijkere maat met meer aandacht en meer vertrouwen.

Het benaderen van bijstandsgerechtigden met positieve aandacht werkt goed in de praktijk, zowel voor werkcoaches en bijstandsgerechtigden, zo blijkt uit kwalitatief onderzoek bij de experimenten. Ondanks het magere resultaat wat de 'harde' uitkomstmaten betreft (zie ook De Boer et al., 2020) schetsen de meeste betrokkenen bij de experimenten (deelnemers, dienstverleners en beleidsmedewerkers) in gesprekken een positief beeld over de experimenten. Zij spreken dan over wat het doet met de deelnemers en dienstverleners en de effectiviteit in de uitvoering. Hoewel niet duidelijk zichtbaar in de cijfers, hebben betrokkenden de indruk dat mensen sneller in beweging komen, zelfredzamer worden, minder trajecten nodig hebben, de angst voor de gemeente is minder en daardoor zijn ook weer andere perspectieven mogelijk. In verschillende gemeenten gaan de dienstverleners met meer plezier naar het werk, ze ervaren meer voldoening en ontdekken dat de relatie met de deelnemer beter wordt waardoor meer dingen mogelijk zijn en de betere verstandhouding ook leidt tot meer beweging en initiatief.

Ook onderstrepen onze resultaten de diversiteit tussen gemeenten in Nederland. Dit ondersteunt een pleidooi voor meer ruimte en middelen voor lokale uitvoerders en politici om te experimenteren en leren in de uitvoering. Door goed, bij voorkeur centraal goed gecoördineerd onderzoek op lokaal niveau te koppelen met ruimte in de wettelijke kaders om op basis van die onderzoeken het beleid op de lokale noden te kunnen aanpassen, kan een effectievere en efficiëntere uitvoering worden bereikt. Tenslotte lijkt het ons raadzaam om, nu door de COVID-19 pandemie een grote instroom van bijstandsgerechtigden met zeer diverse achtergrond en bemiddelbaarheid aanstaande is, de uitvoerende professionals het vertrouwen, de ruimte en de middelen te geven om die instroom efficiënt en effectief weer naar werk te begeleiden. Dat bijstandsgerechtigden en werkcoaches daarbij meer zelfvertrouwen krijgen en de grotere autonomie als prettig ervaren, is belangrijke bijvangst.

Referenties

De Boer, H-W, Bolhaar, J., Jongen, E., & Zulkarnain (2020). Evaluatie experimenten Participatiewet: Effecten op de uitstroom naar werk. CPB: Den Haag

Groot, L., Muffels, R. J., & Verlaat, T. (2019). Welfare states' social investment strategies and the emergence of Dutch experiments on a minimum income guarantee, *Social Policy & Society* (2019)18:2, 277–287, Cambridge University Press

Headey, B., Muffels, R.J.A., & Wagner, G. (2010). Long-running German panel survey shows that personal and economic choices, not just genes, matter for happiness. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 107(42), 17922-17926

Helliwell, J. F., & Huang, H. (2014). New measures of the costs of unemployment: Evidence from the subjective well-being of 3.3 million Americans. *Economic Inquiry*, 52(4), 1485-1502.

Kunzmann, U., Little, T. D., & Smith, J. (2000). Is age-related stability of subjective well-being a paradox? Cross-sectional and longitudinal evidence from the Berlin Aging Study. *Psychology and aging*, 15(3), 511.

Ministerie van Sociale Zaken en Werkgelegenheid (2017). Tijdelijke regeling experimenten Participatiewet. *Staatsblad van het Koninkrijk der Nederlanden*, 69, 1-18. Beschikbaar op: <https://wetten.overheid.nl/BWBR0039258/2019-10-01>

Uslaner, E. M. (2002). *The moral foundations of trust*. Cambridge University Press.