



## UvA-DARE (Digital Academic Repository)

### Het besloten club- en het ingezetenen criterium voor coffeeshops

*Een natuurlijk experiment*

van Ooyen-Houben, M.; Bieleman, B.; Korf, D.; De Witte, K.

**DOI**

[10.5553/TvC/0165182X2017059102002](https://doi.org/10.5553/TvC/0165182X2017059102002)

**Publication date**

2017

**Document Version**

Final published version

**Published in**

Tijdschrift voor Criminologie

**License**

CC BY-NC

[Link to publication](#)

**Citation for published version (APA):**

van Ooyen-Houben, M., Bieleman, B., Korf, D., & De Witte, K. (2017). Het besloten club- en het ingezetenen criterium voor coffeeshops: Een natuurlijk experiment. *Tijdschrift voor Criminologie*, 59(1-2), 10-29. <https://doi.org/10.5553/TvC/0165182X2017059102002>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

## ARTIKELEN

# Het besloten club- en het ingezetenen criterium voor coffeeshops

## Een natuurlijk experiment

*Marianne van Ooyen-Houben, Bert Bieleman, Dirk Korf & Kristof De Witte*

*In Nederland wordt de verkoop van cannabis in coffeeshops gedoogd als voldaan is aan bepaalde criteria. In 2012 kwamen er twee gedoogcriteria bij: het besloten club- en het ingezetenen criterium (de 'wietpas'). Gepland was een fasegewijze invoering: eerst in het zuiden, later in de rest van het land. Dit bood de mogelijkheid tot een natuurlijk experiment. In een experimentele groep van zeven gemeenten in het zuiden en een vergelijkingsgroep in de rest van het land zijn voor- en nametingen verricht van de ervaren overlast rond coffeeshops, het drugstoerisme, het aantal bezoeken aan coffeeshops en de omvang van de illegale gebruikersmarkt. Hierin traden na de implementatie van de nieuwe criteria substantiële veranderingen op in de experimentele groep. Initiële verschillen tussen de twee groepen en uiteenlopende lokale implementatie maakten het echter problematisch te concluderen dat dit effecten waren van de nieuwe criteria. In dit artikel wordt nu onderzocht of de gevonden veranderingen niet toch causaal toegeschreven kunnen worden aan de nieuwe criteria. De conclusie is dat de onderzoekspopzet dat ondanks de methodologische manco's toelaat en dat de veranderingen werden veroorzaakt door de nieuwe criteria.*

## Inleiding

### *Nieuwe criteria voor coffeeshops*

In 2012 werd het beleid voor de Nederlandse coffeeshops aangescherpt. Coffeeshops mochten uitsluitend nog toegang verlenen en verkopen aan leden. De lidmaatschappen moesten gedocumenteerd worden in een controleerbare ledenlijst. Het lidmaatschap was uitsluitend toegankelijk voor ingezetenen van Nederland. Het beleid kreeg vorm in twee nieuwe landelijke gedoogcriteria: het besloten club- en het ingezetenen criterium, die al snel werden betiteld als 'de wietpas'.

Coffeeshops zijn een typisch Nederlands verschijnsel. Het zijn commerciële horecagelegenheden waar cannabis wordt verkocht aan gebruikers. Hoewel dit is verboden in artikel 3 van de Opiumwet, laten de politie en het Openbaar Ministerie de coffeeshophouder ongemoeid. De verkoop wordt gedoogd, hetgeen gedefinieerd wordt als 'het afzien van handhaven van wetten en wettelijke normen, terwijl dat juridisch en feitelijk wel mogelijk is' (TK 30050-2, 2005). Bij de coffeeshops gaat het om 'actief' of 'hard' gedogen: de overheid en het Openbaar Ministerie hebben expliciet te kennen gegeven onder bepaalde condities niet te zullen

handhaven (Projectgroep 'Klem van gedogen', 2004). Die condities – de gedoogcriteria<sup>1</sup> – staan in de Aanwijzing Opiumwet. Het Openbaar Ministerie maakt bij het gedogen gebruik van het opportunitiebeginsel, de bevoegdheid om af te zien van handhaving omdat daarmee een hoger identificeerbaar algemeen belang is gediend. Bij coffeeshops is dit belang erin gelegen dat zij cannabisgebruikers weghouden van de illegale criminele markt, waar vaker naast cannabis (een softdrug) ook drugs met een onaanvaardbaar risico voor de volksgezondheid (harddrugs) te koop zijn. Coffeeshops dragen zo bij aan een 'scheiding der markten' van soft- en harddrugs en dienen het belang van de volksgezondheid. De gereguleerde verkoop in de veilige omgeving van de coffeeshops in plaats van op illegale criminele markten dient tevens de openbare orde.

Rond het coffeeshopsysteem zijn in de loop van de tijd steeds meer problemen gesignaleerd (Van Laar e.a., 2009). Er was, met name in grenssteden, overlast door de vele bezoekers uit het buitenland. Coffeeshops waren volgens de overheid te groot en niet goed beheersbaar en er was (georganiseerde) criminaliteit in relatie tot coffeeshops en drugshandel. Het coffeeshopsysteem stuit bovendien op kritiek in binnen- en buitenland (TK 24077-259, 2011).

Door van coffeeshops besloten clubs te maken voor alleen Nederlandse ingezetenen wilde de overheid de problemen terugdringen en coffeeshops terugbrengen naar kleinschalige beheersbare voorzieningen, zoals ze oorspronkelijk bedoeld waren.

In de discussie over het nieuwe beleid werd gewaarschuwd voor verplaatsing van de cannabisverkoop naar de illegale straathandel (TK Handelingen 58-7 en 69-8, 2012; TK 24077-280, 2012). De minister van Veiligheid en Justitie zegde daar tijdelijk extra handhavingscapaciteit bij de politie voor toe.

### *Evaluatieonderzoek*

De minister zegde de Tweede Kamer onderzoek toe:

'De lokale, regionale en landelijke effecten en de eventuele neveneffecten (zoals een mogelijke toename van illegale straathandel) van de maatregelen zullen uiteraard scherp worden gevolgd.' (TK 24077-259, 2011, p. 4)

Het evaluatieonderzoek is begin 2012 in gang gezet en uitgevoerd door het WODC, het Bonger Instituut en Bureau Intraval. Er zijn twee rapportages verschenen (Van Ooyen-Houben e.a., 2013; 2014). Het onderzoek evalueerde hoe het besloten club- en het ingezetenen criterium uitpakten:

- Hoe zijn de criteria geïmplementeerd?
- Hebben de nieuwe criteria tot de beoogde uitkomsten geleid: minder ervaren overlast, minder drugstoerisme, kleinschaliger en beter beheersbare coffeeshops?

1 Tot 2012 waren dit de AHOJG-criteria: geen Affichering, geen Harddrugs, geen Overlast, geen Jeugdigen, geen Grote hoeveelheden.

Marianne van Ooyen-Houben, Bert Bieleman, Dirk Korf & Kristof De Witte

- Hebben de nieuwe criteria geleid tot een groei van de illegale gebruikersmarkt van cannabis?

In het onderzoek werden substantiële veranderingen waargenomen in de aard van de overlast en de omvang van het drugstoerisme, het coffeeshopbezoek en de illegale gebruikersmarkt na de implementatie van de nieuwe criteria. Hoewel het onderzoek was opgezet als een effectstudie, rezen tijdens de uitvoering twijfels over de mogelijkheden tot causale attributie van deze veranderingen aan de nieuwe criteria, zodat de onderzoekers daar destijds terughoudend in waren.

Dit causaliteitsprobleem vormt de aanleiding voor dit artikel. De kernvraag is: kunnen met de gekozen opzet van het evaluatieonderzoek – en wat daarvan in de praktijk terecht is gekomen – causale conclusies worden getrokken?

Dit artikel is als volgt opgebouwd. Eerst beschrijven we hoe het experiment eruitzag. Daarna geven we de opzet en de belangrijkste bevindingen van het evaluatieonderzoek weer. Dan volgt een kritische beschouwing van het onderzoek, in het bijzonder wat betreft de mogelijkheid tot het doen van causale uitspraken. Deze beschouwing laat zien dat ondanks methodologische manco's toch conclusies over de *oorzaak* van de veranderingen kunnen worden getrokken.

## De nieuwe criteria en de interventielogica

Het evaluatieonderzoek is gestart met een beschrijving van de nieuwe criteria: wat houden ze in, wat is de bedoeling? Hierbij is het conceptuele kader van een 'realist evaluation' gevolgd (Pawson & Tilley, 1997; Pawson, 2013). De 'realist' benadering gaat ervan uit dat elk beleid, impliciet dan wel expliciet, bepaalde veronderstellingen heeft over wat zou moeten worden gedaan en welk doel daarmee bereikt zou moeten worden. Het geheel van deze veronderstellingen is de 'beleids-theorie', of ook, omdat beleid bestaat uit interventies en programma's, de 'interventietheorie' of 'programmatheorie'.

Pawson en Tilley (1997) en Pawson (2013) formuleren een beleidstheorie in termen van een C-M-O: een samenspel van Context, Mechanismen en uitkomsten (Outcomes). Elke beleidsinterventie vindt plaats binnen een bepaalde context (bij een bepaalde doelgroep, voor een bepaald probleem), waarin (gedrags)veranderingen (mechanismen) *getriggerd* moeten worden, die dan moeten leiden tot de beoogde uitkomsten.

Het blootleggen van (psychologische) mechanismen is een intensieve bezigheid omdat ze vaak impliciet zijn. Er wordt in evaluaties dan ook regelmatig gekozen voor een meer pragmatische benadering, waarbij niet zozeer de impliciete veronderstellingen over werkzame mechanismen, als wel de veronderstelde processen op een meer globaal niveau worden beschreven. Het gaat dan om een 'logica' in plaats van een 'theorie'. Hier is in het onderhavige onderzoek voor gekozen. De term 'interventielogica' drukt uit dat het plannen waren uit overleg tussen ministerie, gemeenten en politie (niet alleen 'beleid').

Voor het reconstrueren van de interventielogica hebben de onderzoekers landelijke, (boven)regionale en lokale beleidsplannen uit 2008-2012 geanalyseerd. De conceptlogica is bij T0 ter validering voorgelegd aan 40 actoren op nationaal, regionaal en lokaal niveau.

De interventielogica van het besloten club- en het ingezetenen criterium (figuur 1) laat zien wat de aannames waren over de problematiek en binnen welke juridische en beleidskaders de oplossing gezocht moest worden (Context). Het experiment met het besloten club- en het ingezetenen criterium was een complex proces van communicatie, informatie, overleg, afstemming, controle, handhaving en naleving. Van verschillende actoren – het ministerie van Veiligheid en Justitie en het College van procureurs-generaal, de burgemeesters, het Openbaar Ministerie en de politie, en de coffeeshophouders en hun klanten – werd bepaald gedrag verwacht (Mechanismen). Als ze allemaal de beoogde bijdrage leveren, zullen de nieuwe criteria, zo was de veronderstelling, tot minder overlast, minder drugstoerisme, kleinere en beheersbare coffeeshops, minder criminaliteit in relatie tot coffeeshops en drugshandel en minder kritiek op het coffeeshopsysteem leiden en kunnen eventuele negatieve neveneffecten worden tegengegaan (Outcomes).

De interventielogica was de inhoudelijke leidraad voor de evaluatie. Onderzocht is of en hoe de activiteiten en uitkomsten zijn gerealiseerd. Omdat de logica de plannen expliciteert en ordent, kan een interventie er systematisch en volledig mee geëvalueerd worden.

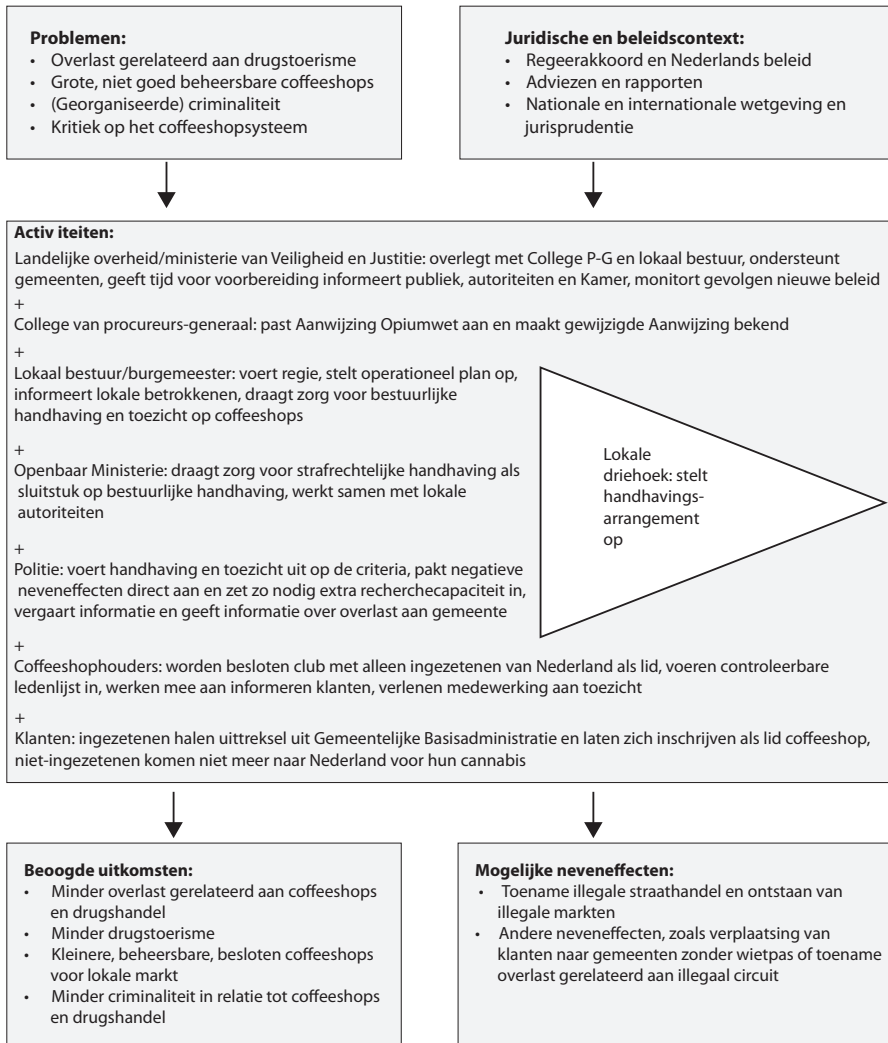
## De opzet van het onderzoek

### *Een natuurlijk experiment*

De daadwerkelijke implementatie van de nieuwe criteria was gefaseerd gepland. Per 1 mei 2012 startten Limburg, Noord-Brabant en Zeeland. In de 23 coffeeshopgemeenten daar moesten de beoogde activiteiten uitgevoerd worden (zie figuur 1). In de coffeeshopgemeenten in de rest van het land gebeurde tot 1 januari 2013 niets. Vanaf 2013 zouden de nieuwe criteria ook daar geïmplementeerd worden.

Door deze gefaseerde uitrol ontstond een situatie waarin een externe actor (niet de onderzoekers, maar beleidsmakers) een conditie mét en een conditie zonder interventie creëerde. Op die manier ontvouwde zich een situatie met een experimentele groep in het zuiden en een vergelijkingsgroep in de rest van het land, een 'natuurlijk experiment' (Swanborn, 2005; Murnane & Willett, 2011). Dit bood de gelegenheid tot het meten van de effecten van het nieuwe beleid. Er is sprake van een *discontinuity design*: door een beleidsverandering werd van 1 mei 2012 tot januari 2013 een grens getrokken midden door het land. Daarna zouden de beide criteria over het hele land uitgerold worden. In de vergelijkingsgroep was aldus sprake van *delayed treatment*.

Marianne van Ooyen-Houben, Bert Bieleman, Dirk Korf & Kristof De Witte



**Figuur 1** De interventielogica van het besloten club- en het ingezetenen criterium

*De steekproef*

De steekproef is getrapd samengesteld en bestaat uit eenheden op verschillende niveaus: gemeenten, coffeeshops, omwonenden en ondernemers in de directe nabijheid van coffeeshops, klanten van coffeeshops en actuele gebruikers van cannabis buiten de coffeeshops. De (sub)steekproeven zijn zo representatief mogelijk samengesteld door rekening te houden met bekende kenmerken, wervingsmomenten evenwichtig te spreiden, en respondenten aselekt of op verschillende manieren en tijdstippen te benaderen:

**Tabel 1** Aantal coffeeshopgemeenten in de steekproef, onderverdeeld naar experimentele groep en vergelijkingsgroep en categorie

	Experimentele groep	Vergelijkingsgroep
Grensgemeenten met veel softdrugstoerisme	2	2
Grensgemeenten met weinig softdrugstoerisme	1	1
Niet-grensgemeenten met veel softdrugstoerisme	1	1
Niet-grensgemeenten met weinig softdrugstoerisme	3	3
Totaal	7	7

- Uit de 104 Nederlandse gemeenten met coffeeshops (eind 2011; Bieleman e.a., 2012) zijn veertien gemeenten geselecteerd. De selectie geschiedde in overleg tussen de onderzoekers en beleidsmedewerkers van het ministerie van Veiligheid en Justitie. Zeven gemeenten zijn gekozen uit de coffeeshopgemeenten in de drie zuidelijke provincies. Geselecteerd zijn grotere, middelgrote en kleinere gemeenten, gemeenten met veel en gemeenten met weinig coffeeshops, gemeenten met veel en gemeenten met weinig (soft)drugstoerisme en zowel grensgemeenten als gemeenten verder weg van de grens (tabel 1).
- Deze gemeenten zijn gematcht met zeven gemeenten in de rest van het land op de variabelen: aantal inwoners, aantal coffeeshops, mate van aanwezigheid van softdrugstoerisme – dit is afgeleid uit voorhanden onderzoeken – en geografische ligging ten opzichte van de landsgrens. Zowel in de vergelijkingsgroep als in de experimentele groep bevonden zich gemeenten met veel en weinig drugstoerisme en veel en weinig overlast en toeloop naar de coffeeshops. Verder is gestreefd naar een evenwichtige geografische spreiding van de gehele steekproef. De onderliggende assumptie van de matching was dat de gemeenten in de experimentele groep zonder de invoering van het nieuwe beleid gelijkaardige ontwikkelingen zouden laten zien als de gematchte gemeenten in de vergelijkingsgroep. De gemeenten in de vergelijkingsgroep laten zodoende toe om de ‘counterfactual’ te bepalen: de niet-geobserveerde uitkomst in de experimentele groep indien geen interventie had plaatsgevonden.

Kort voor ons onderzoek liep ander onderzoek naar overlast rond coffeeshops in elf gemeenten. Deze gemeenten zijn uitgesloten van deelname, op één gemeente na die vanwege de omvang van de problematiek te grote relevantie had voor ons onderzoek. Vlak buiten de drie zuidelijke provincies gelegen gemeenten (zoals Nijmegen of Dordrecht), die mogelijk *catchment area* (‘overloopgebied’) konden worden voor softdrugstoeristen (Bowers e.a., 2011; Braga e.a., 2012), zijn uitgesloten van de vergelijkingsgroep. In deze gemeenten kon mogelijk contaminatie

Marianne van Ooyen-Houben, Bert Bieleman, Dirk Korf & Kristof De Witte

plaatsvinden doordat cannabisgebruikers uitwijken naar coffeeshops net over de experimentele grens. Dit was wel interessant geweest met het oog op het onderzoeken van verplaatsingseffecten, maar het had het onderzoek complexer en uitgebreider gemaakt.

De burgemeesters zijn per brief verzocht om medewerking aan het onderzoek; (slechts) één weigerde.

Binnen elke gemeente hebben de onderzoekers in overleg met de coffeeshopambtenaar een coffeeshopgebied geselecteerd met een tot vier coffeeshops, in totaal 32. Uitgesloten zijn gebieden waar per 1 januari 2014 ook het afstandscriterium zou worden ingevoerd, omdat anticipatie hierop tot vertekening in de resultaten zou kunnen leiden.

### *De metingen*

In beide groepen zijn pre- en postmetingen uitgevoerd: een nulmeting (T0) in maart-april 2012, een eerste vervolgmeting (T1) in oktober-november 2012 en een tweede vervolgmeting in oktober-november 2013 (T2).

Naast de uitkomstmetingen is ook een procesevaluatie uitgevoerd. In interviews met in totaal 137 *stakeholders* is nagegaan of en in hoeverre de nieuwe criteria daadwerkelijk volgens plan werden geïmplementeerd. Gestreefd is naar een vertegenwoordiging van de vier partijen die volgens de interventielogica een belangrijke rol hebben: gemeente, politie, Openbaar Ministerie en coffeeshopexploitanten. Dit is in de meeste gemeenten gelukt. Ook zijn landelijke en regionale actoren geïnterviewd. In het experimentele gebied zijn interviewrondes gehouden bij T1 en T2, in het vergelijkingsgebied bij T2.

De afhankelijke variabelen zijn gemeten bij verschillende categorieën respondenten:

- Om de ervaren overlast te meten zijn alle omwonenden en ondernemers binnen een straal van maximaal 100 meter rond de coffeeshops telefonisch benaderd voor deelname aan een enquête. De telefoonnummers zijn verkregen van aanbieders van adresgegevens en uit het Handelsregister van de Kamer van Koophandel. Daarnaast hebben onderzoekers aangebeld en bewoners gevraagd mee te werken en een telefoonnummer af te geven. De omwonenden zijn vervolgens op verschillende dagen en tijdstippen geënquêteerd. In één coffeeshopgebied was binnen de gegeven straal geen enkele omwonende. De respons was 63 procent bij T0, 59 procent bij T1 en 66 procent bij T2. Dit resulteerde in 712 respondenten bij T0, 714 bij T1 en 714 bij T2.
- Om de omvang van coffeeshopbezoeken te meten zijn bij de coffeeshops tellingen uitgevoerd van bezoeken, ten minste 40 per coffeeshop, op verschillende dagen en tijdstippen: 1.320 bij T0, 1.641 bij T1 en 1.567 bij T2. In de experimentele groep waren bij de nametingen enkele coffeeshops gesloten. Omdat deze uitval mogelijk niet random was, kan een zekere mate van selectie zijn ontstaan.
- Om de omvang van het drugstoerisme in coffeeshops en aanschafgedrag in coffeeshops dan wel op de illegale markt te meten zijn bij bezoekers van de coffeeshops bij binnenkomst face-to-face vragenlijsten (in het Nederlands,



Duits, Frans of Engels) afgenomen over hun ingezetenschap en aanschafgedrag: 1.051 bij T0, 739 bij T1 en 726 bij T2. De bezoekers werden aselekt benaderd op verschillende dagen en tijdstippen. Naast deze crosssectionele metingen is ook een cohort van 108 bezoekers gevolgd.

- Dezelfde afhankelijke variabelen zijn gemeten in face-to-face straatenuêtes onder actuele cannabisgebruikers die hun cannabis hadden aangekocht in de gemeente van onderzoek. Dit gebeurde op plekken (zoals winkelgebieden, uitgaanspleinen, parken en stations) met een zo goed mogelijke dwarsdoorsnede van cannabisgebruikers onder voorbijgangers en per gemeente steeds verspreid over drie dagen (donderdag, vrijdag, zaterdag), zowel 's middags als 's avonds. Gevraagd werd naar het aanschafgedrag, meer specifiek: kopen ze hun cannabis in de coffeeshop of op de illegale markt? De enquête werd niet afgenomen in de directe nabijheid van coffeeshops, aangezien dit vrijwel zeker een vertekend beeld zou opleveren van het koopgedrag. In totaal ging het om 942 (T0), 812 (T1) en 907 (T2) gebruikers, met aanvullend een cohort van 79 gebruikers.
- In vier gemeenten is etnografisch veldonderzoek uitgevoerd ter validering en verdieping van de resultaten uit de andere metingen en met nadruk op het verkrijgen van inzicht in verschuivingen en kenmerken van de illegale gebruikersmarkt.

De minister wilde met de nieuwe criteria ook de (georganiseerde) criminaliteit in relatie tot coffeeshops en drugshandel – meer specifiek: de hennep teelt – terugdringen en tegemoetkomen aan de (buitenlandse) kritiek op het coffeeshopsysteem. Om dit op valide wijze te meten zou apart onderzoek nodig zijn. Deze uitkomstvariabelen zijn daarom buiten beschouwing gelaten.

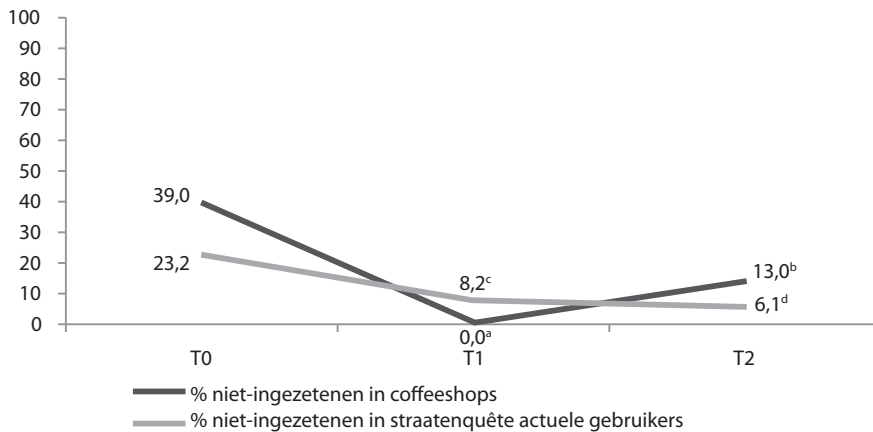
### **Belangrijkste bevindingen**

In het experimentele gebied deden zich na mei 2012 substantiële veranderingen voor op de gebruikersmarkt van cannabis. De drugstoeristen bleven grotendeels weg, het aantal bezoeken aan coffeeshops daalde fors en gebruikers schaften hun cannabis aanzienlijk vaker aan via illegale verkoopkanalen, zo bleek uit de metingen bij T1 (figuren 2, 3 en 4). De omvang en frequentie van ervaren overlast veranderden weinig, maar in de aard van de overlast trad een verschuiving op: vóór de nieuwe criteria ervoeren omwonenden met name overlast van coffeeshops, daarna hadden ze vooral overlast van dealactiviteiten op straat. Deze trends werden eveneens waargenomen door de geïnterviewde lokale experts.

De veranderingen deden zich al snel na de handhaving van de nieuwe criteria voor en werden minder waargenomen in de vergelijkingsgroep. Er waren daar geen significante veranderingen in coffeeshopbezoek, de mate van ervaren overlast van coffeeshops of drugstoeristen en de aanwezigheid van niet-ingezetenen en minder veranderingen in de aanschaf van cannabis buiten de coffeeshop.

Het besloten-clubcriterium pakte niet goed uit. Veel ingezetenen lieten zich niet als lid van een coffeeshop registreren en weken (net als een deel van de drugstoe-

Marianne van Ooyen-Houben, Bert Bieleman, Dirk Korf & Kristof De Witte



<sup>a</sup>T1-T0: Pearson Chi-Square=180,76, df=1, p=,00

<sup>b</sup>T2-T1: Pearson Chi-Square=47,43, df=1, p=,00

<sup>c</sup>T1-T0: Pearson Chi-Square=29,742, df=1, p=,00

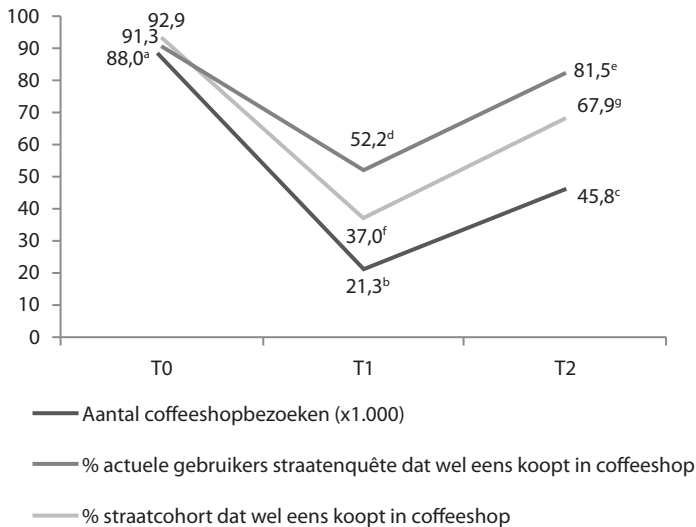
<sup>d</sup>T2-T1: Pearson Chi-Square=1,052, df=1, p=,31

**Figuur 2** Niet-ingezetenen volgens enquête coffeeshopbezoekers en straatenquête actuele gebruikers, drie zuidelijke provincies, T0, T1 en T2

risten) uit naar de illegale markt. De nieuwe criteria zaagden zo aan de stoelpoten van het gedoogbeleid. Nog tijdens het onderzoek, in november 2012, paste de minister de interventie ingrijpend aan: het besloten-clubcriterium verviel en de beslissing tot daadwerkelijke handhaving van het ingezetenen criterium werd bij de lokale autoriteiten gelegd (TK 24077-293, 2012).

Eind 2013, bij T2, bleek dat na deze afzwakking van de interventie ook de uitkomsten verflauwden, zij het dat niet alle verschuivingen tussen T1 en T2 even robuust waren als die tussen T0 en T1 (figuren 2, 3 en 4). De coffeeshops in het zuiden werden weer voller; ingezetenen kwamen (deels) terug. Niet-ingezetenen mochten op veel plaatsen ook gewoon (weer of nog) gebruik maken van de coffeeshops. Ze waren echter bij T2 nog grotendeels afwezig, niet alleen in de coffeeshops, maar ook daarbuiten. De meeste gemeenten handhaafden het ingezetenen criterium niet actief – op de meeste plaatsen was simpelweg nauwelijks overlast of softdrugstoerisme en waren de coffeeshops goed beheersbaar. In het zuiden werd in enkele gemeenten nog gehandhaafd, in het noorden nergens. De illegale markt nam af. Omwonenden en ondernemers in de directe omgeving van coffeeshops ervoeren wel weer meer overlast gerelateerd aan coffeeshops.

Zo schoof in de experimentele groep in 2013 alles weer op in de richting van de situatie van vóór de invoering van de nieuwe criteria, maar niet voor 100 procent (figuren 2, 3 en 4). In de vergelijkingsgroep waren de veranderingen beperkt, daar bleef het relatief rustig.



<sup>a</sup> T0: 95% betrouwbaarheidsmarge=75.838-100.226, gemiddeld 88.032

<sup>b</sup> T1: 95% betrouwbaarheidsmarge=17.088-25.547, gemiddeld 21.312; T1-T0: significant verschil (marges overlappen niet, zie bijlage tabel A), p=,00

<sup>c</sup> T2: 95% betrouwbaarheidsmarge=40.261-51.233, gemiddeld 45.747; T2-T1: significant verschil (marges overlappen niet, zie bijlage tabel A), p=,00

<sup>d</sup> T1-T0: Pearson Chi-Square=148,561, df=1, p=,00

<sup>e</sup> T2-T1: Pearson Chi-Square=68,074, df=1, p=,00

<sup>f</sup> T1-T0: Related samples McNemar Test=13,067, df=1, p=,00

<sup>g</sup> T2-T1: Related samples McNemar Test=4,267, df=1, p=,04

**Figuur 3** *Aanschaf cannabis in coffeeshops volgens tellingen bezoeken, straatenquête actuele gebruikers en straatcohort gebruikers (ingezetenen, drie zuidelijke provincies, T0, T1 en T2)*

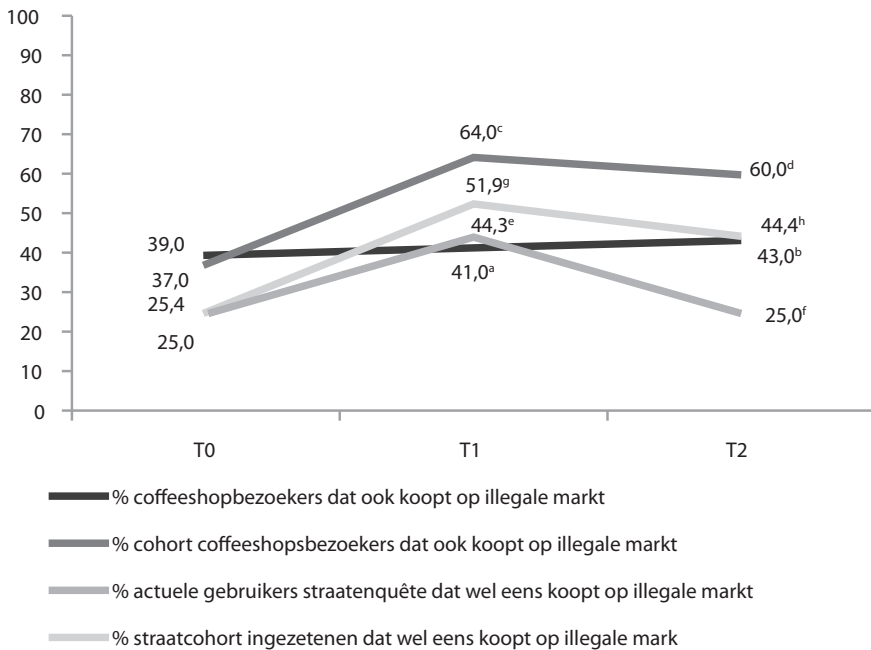
### Kritische beschouwing van de opzet in het licht van causaliteit

De vraag die in dit artikel voorligt, is of de gevonden veranderingen tussen T0, T1 en T2 causaal toegeschreven kunnen worden aan het besloten club- en het ingezetenen criterium: waren het *effecten* van de nieuwe criteria?

#### *Voorwaarden voor het aantonen van causaliteit*

Idealiter worden effecten gemeten met een gerandomiseerd experiment (Swanborn, 2005; Welsh & Farrington, 2007; Murnane & Willett, 2011). Bij toevalstoeuwijzing door de onderzoeker kan ervan uitgegaan worden dat de experimentele en de vergelijkingsgroep een gelijke startsituatie hebben op alle relevante variabelen, zowel geobserveerde als niet-geobserveerde. Waargenomen veranderingen in uitkomsten kunnen dan valide geattribueerd worden aan de interventie.

Marianne van Ooyen-Houben, Bert Bieleman, Dirk Korf &amp; Kristof De Witte



<sup>a</sup> T1-T0: Pearson Chi-Square=0,16, df=1, p=,69

<sup>b</sup> T2-T1: Pearson Chi-Square=0,59, df=1, p=,44

<sup>c</sup> T1-T0: Related samples McNemar Test=6,50, df=1, p=,01

<sup>d</sup> T2-T1: Related samples McNemar Test=1,79, df=1, p=,18

<sup>e</sup> T1-T0: Pearson Chi-Square=29,710, df=1, p=,00

<sup>f</sup> T2-T1: Pearson Chi-Square=28,815, df=1, p=,00

<sup>g</sup> T1-T0: Related samples McNemar Test=4,00, df=1, p=,04

<sup>h</sup> T2-T1: Related samples McNemar Test=0,00, df=1, p= 1,00

**Figuur 4** *Aanschaf cannabis buiten de coffeeshops (illegale markt) volgens enquête coffeeshopbezoekers (crosssectioneel en cohort), straatenquête actuele gebruikers en straatcohort gebruikers (ingezetenen), drie zuidelijke provincies, T0, T1 en T2*

In dit onderzoek konden de onderzoekers niet randomiseren. Het was een politiek besluit om gemeenten in het zuiden te laten starten met de nieuwe criteria en de andere gemeenten nog niet. Dit creëerde wel de mogelijkheid van een natuurlijk experiment.

Uit natuurlijke experimenten kunnen evenzeer causale conclusies worden getrokken als uit gerandomiseerde experimenten, maar ze moeten dan wel aan bepaalde voorwaarden voldoen. Ze zijn in termen van causaliteit op hun best wanneer proefpersonen op basis van een loterij al dan niet worden toegewezen aan een experimentele of een vergelijkingsconditie. Dit is bijvoorbeeld het geval bij cursussen of interventies waar de vraag het aanbod overstijgt en mensen toevallig buiten de boot vallen. Dan is er, ondanks het feit dat de onderzoeker niet zelf zijn groepen kan samenstellen, toch sprake van een opzet met random toewijzing.

Dergelijke situaties doen zich echter (ook in de criminologie) niet zo vaak voor (Murnane & Willett, 2011). Dat geldt ook voor het onderhavige onderzoek.

De cruciale vraag is nu in hoeverre de gekozen opzet causale gevolgtrekkingen toelaat. We toetsen onze opzet aan de voorwaarden waaraan dan voldaan moet zijn. Volgens Murnane en Willett (2011), die uitgebreid ingaan op natuurlijke experimenten, zijn dit de volgende:

- De respondenten maken deel uit van een gedefinieerde populatie.
- De experimentele condities moeten expliciet en helder gedefinieerd zijn.
- Van de gemeten uitkomsten kan verwacht worden dat ze gevoelig zijn voor de gevolgen van de interventie.
- De respondenten hebben geen invloed op de toewijzing aan de condities en zijn niet geselecteerd op kenmerken die samenhangen met de uitkomstmaat.
- Ze zijn *equal in expectation initially*: allerlei niet-geobserveerde variabelen zijn niet ongelijk verdeeld over de groepen.

*Is aan de voorwaarden voldaan?*

Maken de respondenten deel uit van een gedefinieerde populatie? Aan deze eerste voorwaarde lijkt te zijn voldaan. De populaties van gemeenten, coffeeshops, bezoekers, omwonenden en actuele gebruikers zijn vooraf gedefinieerd en afgeleid. We weten op welke groepen de conclusies betrekking hebben.

Zijn de experimentele condities expliciet en helder gedefinieerd? Ook deze voorwaarde lijkt in orde. De gemeenten in de experimentele groep bereidden zich gezamenlijk voor op het nieuwe beleid. Ze pakten zich samen in de 'brede ambtelijke werkgroep zuid' en werkten de nieuwe criteria daar concreet en eenduidig uit, zodat ze in de implementatie één lijn konden trekken.

Kan van de gemeten uitkomsten verwacht worden dat ze gevoelig zijn voor de gevolgen van de interventie? De onderzoekers hebben bij het opstellen van de interventielogica uitgebreide analyses uitgevoerd op de beleidsplannen, mede om de verwachte gevolgen goed in beeld te krijgen (zie figuur 1). Ook is een literatuurstudie uitgevoerd naar eerdere interventies in het Nederlandse coffeeshopsysteem en – internationaal – in de cannabisconsumentenmarkt. Deze onderschreef dat verwacht mocht worden dat het drugstoerisme en de daarmee samenhangende overlast zouden afnemen als gevolg van de nieuwe maatregelen en dat er een zekere verplaatsing van de cannabisverkoop kon optreden van coffeeshops naar de illegale markt (Van Ooyen-Houben e.a., 2014). Ook vanuit de theorie werden de verwachtingen ondersteund. De gelegenheidstheorie voorspelt dat als de toegang tot de coffeeshops ('aantrekkelijke doelwitten') wordt geblokkeerd, voor niet-ingezetenen ('gemotiveerde daders') de prikkel wegvalt om hiervoor naar Nederland te komen en ze ervoor zullen kiezen weg te blijven. De verplaatsingstheorie voorspelt een 'waterbedeffect' van de cannabisverkoop, zij het niet zozeer naar omliggende gebieden, maar eerder naar verderaf gelegen gebieden (voor drugstoeristen bijvoorbeeld naar de markt in eigen land, waar het nieuwe beleid niet van toepassing is). Aan deze voorwaarde lijkt voldaan.

Marianne van Ooyen-Houben, Bert Bieleman, Dirk Korf & Kristof De Witte

Hebben de respondenten geen invloed op de toewijzing aan de condities, zijn ze niet geselecteerd op kenmerken die samenhangen met de uitkomstmaat en zijn ze ‘equal in expectation initially’?

Dit zijn de belangrijkste en tegelijkertijd de meest problematische voorwaarden (de laatste twee). Ze betreffen de vergelijkbaarheid van de groepen. Kenmerken van de onderzoekseenheden die de uitkomstvariabelen kunnen beïnvloeden, moeten random verdeeld zijn over de groepen, respondenten mogen zelf geen invloed hebben op de selectie en er mag niet op een relevante variabele geselecteerd zijn. We moeten kunnen aannemen dat allerlei niet-geobserveerde variabelen niet ongelijk verdeeld zijn over de groepen.

Dit ligt lastig bij ons design. Gemeenten waren betrokken bij de besluitvorming over de inwerkingtreding van de nieuwe criteria en in diverse beleidsdocumenten is te lezen dat het zuiden van het land als eerste startte omdat beleidsmakers dáár wilden starten waar de problemen het meest acuut waren. Gemeenten, coffeeshops, omwonenden, coffeeshopklanten en cannabisgebruikers in de drie zuidelijke provincies hadden hiermee een grotere kans om in de experimentele groep terecht te komen dan hun evenknieën in de rest van het land. Er bestaat een kans dat dit ook verschillen in andere (niet-)geobserveerde kenmerken en trends tussen de gemeenten in de experimentele en vergelijkingsgroep met zich meebrengt (Swanborn, 2005).

We hebben deze bedreigingen van de validiteit vooraf op verschillende manieren proberen te ondervangen:

- Ten eerste door de *matching* van de gemeenten in de experimentele en vergelijkingsgroep. Deze heeft plaatsgevonden op een aantal bekende variabelen die samenhangen met de beoogde uitkomsten (zie tabel 1). De aanname van de matching is dat we door selectie op geobserveerde kenmerken ook een vergelijkbare groep krijgen op de niet-geobserveerde kenmerken en dat zo twee groepen ontstaan van zo vergelijkbaar mogelijke gemeenten. Hoewel de matching de belangrijkste variabelen zou moeten dekken, bleek het lastig om voor bepaalde gemeenten in het zuiden een zo vergelijkbaar mogelijke gemeente in de rest van het land te vinden. Er is bijvoorbeeld niet gematcht op mogelijke *confounders* als media-aandacht voor het nieuwe beleid, die gevoeligheid voor de interventie kan aanwakkeren, of op de bereidheid van burgemeesters om het nieuwe beleid in praktijk te brengen. Dergelijke *confounders* zijn in de kwalitatieve interviews wel nagevraagd, maar niet zodanig gekwantificeerd dat ze als matchingsvariabele of covariaat konden dienen. Evenmin zagen we een mogelijkheid om de groepsindeling op een instrumentele variabele te baseren. Een andere indeling dan zuiden versus rest van het land zou ook beleidsmatig de plank misslaan: Kamer en minister wilden juist weten wat er gebeurde in het zuiden van het land.
- Ten tweede is gekozen voor een *realist evaluation*-benadering. Deze biedt aanvullende mogelijkheden om veranderingen causaal te duiden. Bij de reconstructie van de interventielogica zijn de onderliggende *causal pathways* expliciet gemaakt, en als de interventie dan inderdaad conform deze logica verloopt, versterkt dat een causale attributie van veranderingen aan de interventie. Het kan dan *aannemelijk* gemaakt worden dat de nieuwe maatregelen heb-

ben bijgedragen aan de uitkomsten ('contribution', zie Leeuw, 2012). Dit geldt des te meer als de interventielogica wordt ondersteund door voorhanden onderzoek of verklarende theorieën (Vaessen, 2010). Bij een realist evaluation hoort een meting van de implementatie: wordt de interventie volgens plan uitgevoerd? Dit was hier relevant omdat veel weerstand bestond tegen de nieuwe criteria, zowel bij coffeeshopklanten als bij coffeeshopeigenaren en sommige burgemeesters. Het was nog maar de vraag of de nieuwe criteria wel overal gehandhaafd en nageleefd zouden worden. Als dat niet zo zou zijn, zouden causale conclusies op losse schroeven komen te staan.

- Ten derde is gekozen voor een *multi-method-* (of *mixed methods-*)evaluatie, die in de wetenschappelijke literatuur wordt aangeraden met het oog op de interne en externe validiteit (o.a. Vaessen, 2010). De drie onderzoeksinstituten verzamelden onafhankelijk van elkaar data bij verschillende groepen respondenten en met verschillende meetinstrumenten. Door kwalitatief en kwantitatief onderzoek te combineren kunnen hun voordelen elkaar versterken en hun nadelen geminimaliseerd worden. Als deze aanpak in eensluitende resultaten zou resulteren, zou dat, zo was de aanname, de validiteit van het onderzoek ten goede komen.

Met deze maatregelen hoopten we aan de voorwaarden voor het doen van causale uitspraken te voldoen.

#### *Tegenvallers tijdens het onderzoek*

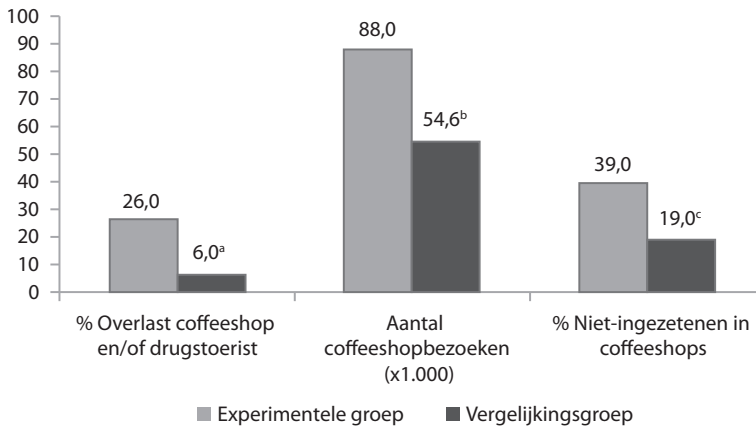
De werkelijkheid gedroeg zich anders dan de onderzoekers voor ogen hadden. Als eerste bleek bij T0 dat, ondanks de matching, in de experimentele groep significant meer overlast werd ervaren van coffeeshops en drugstoeristen en dat er significant meer coffeeshopbezoeken en drugstoeristen waren dan in de vergelijkingsconditie (figuur 5). De matching was blijkbaar niet afdoende. Dit is een van de meest voorkomende problemen bij *discontinuity designs* (Murnane & Willett, 2011). Het ligt voor de hand te veronderstellen dat de groepen dan ook op niet-geobserveerde variabelen verschillen, met als gevolg een dubbele aantasting van de voorwaarden voor het aantonen van causaliteit (Swanborn, 2005).

De selectiebias had geadresseerd kunnen worden door de bandbreedte van de steekproef te verkleinen. We hadden gemeenten met meer afwijkende scores op overlast, drugstoerisme en coffeeshopbezoek buiten de analyses kunnen laten. Dan waren echter gemeenten met de meeste problematiek buiten het onderzoek gevallen. Vanuit oogpunt van beleidsrelevantie was dit geen optie. De grootste belangstelling ging juist uit naar de gemeenten met de meeste problemen en niet naar gemeenten met middelmatige scores op overlast en drugstoerisme. Ook zouden de onderzoeksgroepen kleiner zijn geworden, waardoor het onderzoek aan power had ingeboet. Hier is niet voor gekozen.

Een andere mogelijkheid was om in de analyses een difference-in-differences-benadering te volgen, waarbij gecorrigeerd wordt voor de initiële verschillen, ervan uitgaande dat aan de assumpties hiervoor voldaan is, zoals de aanname van een gelijkaardige trend in de groepen bij afwezigheid van een interventie. Gezien



Marianne van Ooyen-Houben, Bert Bieleman, Dirk Korf &amp; Kristof De Witte



<sup>1</sup> Schatting aantal bezoeken (x1.000) op basis van tellingen door onderzoekers

<sup>a</sup> Experimentete groep-vergelijkingsgroep: Pearson Chi-Square=51,31, df=1, p=,00

<sup>b</sup> Experimentele groep: 95% betrouwbaarheidsmarge=75.838-100.226, gemiddeld 88.032; vergelijkingsgroep: 95% betrouwbaarheidsmarge=44.863-64.311, gemiddeld 54.587; significant verschil (marges overlappen niet, zie bijlage tabel B), p=,00

<sup>c</sup> Experimentele groep-vergelijkingsgroep: Pearson Chi-Square=51,37, df=1, p=,00

**Figuur 5** *Percentage omwonenden en ondernemers binnen 100 meter van een coffeeshop dat overlast ervaart, aantal coffeeshopbezoeken en percentage niet-ingezetenen onder coffeeshopbezoekers, experimentele en vergelijkingsgroep, T0*

de ongelijkheid op initiële (geobserveerde en waarschijnlijk ook niet-geobserveerde) variabelen was deze optie omgeven met de nodige vraagtekens.

Een tweede complicatie betrof de helderheid van de experimentele conditie (de tweede voorwaarde). Bij T1 bleek dat de gemeenten in het zuiden ondanks de gezamenlijke voorbereiding met onduidelijkheden in de uitvoering kampten en dat ze al snel uiteenlopend handhaafden. Sommige gemeenten gingen kordaat en snel van start en deden wat de bedoeling was, andere keken de kat uit de boom en veranderden niet veel. Op de meeste plaatsen daalde het aantal controles al snel na 1 mei en op sommige plaatsen werden de coffeeshops na een maand de facto helemaal niet meer gecontroleerd. De coffeeshops zelf leefden de nieuwe criteria in de meeste gevallen wél na (er was *implementation fidelity*), zij het dat er overtredingen waren (waarvoor ze meestal een waarschuwing kregen), enkele coffeeshops de deuren (al dan niet tijdelijk) sloten en manieren werden gezocht om onder de regels uit te komen (zoals het inrichten van 'neutrale zones' voor niet-leden). De implementatie was zodoende minder strikt dan gepland en de experimentele conditie was omgeven met onduidelijkheid. Qua handhaving liep de scheidslijn tussen experimentele groep en vergelijkingsgroep niet zozeer tussen de drie zuidelijke provincies en de rest van het land, maar tussen de enkele zuidelijke gemeenten die wel handhaafden en de overige gemeenten die dat niet deden.



Vanwege de onduidelijkheid van de interventie en de selectiebias (de tweede en de laatste twee voorwaarden) is besloten om géén (interne validiteit suggererende) berekeningen van effecten en effectgroottes uit te voeren. In plaats van te focussen op verschillen in ontwikkeling tussen de experimentele en vergelijkingsgroep zijn verschillen tussen de metingen berekend voor de experimentele en vergelijkingsgroep afzonderlijk. Zijn causale uitspraken nu niet meer mogelijk?

#### *Bevindingen die een causale uitspraak ondersteunen*

Misschien was het te hoog gegrepen om bij zo'n nieuwe en brede interventie een quasi-experimentele onderzoeksopzet na te streven met veertien gemeenten in het hele land. Maar dankzij de gekozen brede opzet staan we niet met lege handen en beschikken we over informatie die causale interpretaties ondersteunt.

Dankzij de quasi-experimentele opzet met twee groepen werd duidelijk dat in de vergelijkingsgroep nauwelijks significante veranderingen optraden in de uitkomstvariabelen, terwijl zich in dezelfde periode in de experimentele groep robuuste en significante veranderingen voordeden. Dit wijst sterk op een effect van de nieuwe criteria in de experimentele groep in het zuiden. De vergelijkingsgroep bewijst hier zijn nut.

Daar kwam bij dat de betrokken actoren in de interviews zonder uitzondering en met overtuiging aangaven dat de waargenomen veranderingen veroorzaakt werden door de nieuwe criteria.

Bij T2 hadden we te maken met een afgezwakte interventie. Het besloten-clubcriterium bestond niet meer en in de meeste gemeenten, ook de zuidelijke, werd het ingezetenen criterium niet actief gehandhaafd. Voor het onderzoek was dit een *opportunity*, want nu kon bij T2 gemeten worden wat er gebeurde als de interventie (grotendeels) werd weggelaten. Verrassenderwijs werd hierdoor het inzicht in de causale rol van de criteria verscherpt. We zagen bij T2 nog wel de bij T1 geconstateerde veranderingen, maar ook bewegingen terug. Ook dit is een krachtig signaal dat de T1-veranderingen te wijten waren aan de nieuwe criteria.

Een verdere ondersteuning van causaliteit komt uit de multi-method-benadering. Het onderzoek gebruikte meerdere bronnen en meetmethoden en werd uitgevoerd door drie instituten, die tijdens de dataverzameling onafhankelijk van elkaar te werk gingen. De bevindingen uit de verschillende bronnen waren buitengewoon consistent. De tellingen van coffeeshopbezoeken en de enquête onder coffeeshopbezoekers (crosssectioneel en cohort, uitgevoerd door Intraval), de straatenquête onder actuele cannabisgebruikers (crosssectioneel en cohort, uitgevoerd door het Bonger Instituut) en de lokale experts (geïnterviewd door het WODC) leggen dezelfde trends bloot in het coffeeshopbezoek. De enquêtes onder coffeeshopbezoekers en actuele gebruikers op straat laten soortgelijke ontwikkelingen zien in de omvang van het drugstoerisme en op de illegale markt, die bovendien ondersteund werden door lokale experts en etnografisch veldonderzoek. Alleen de enquête onder coffeeshopbezoekers laat een vrij constant aandeel van aanschaf op de illegale markt zien – mogelijk omdat dit een groep vaste klan-

Marianne van Ooyen-Houben, Bert Bieleman, Dirk Korf & Kristof De Witte

ten van coffeeshops betreft. Geen enkele bevinding weerlegt een causale interpretatie. De richting van de ontwikkelingen is eenduidig (zie figuren 2, 3 en 4).

Bovendien – en last but not least – worden onze bevindingen ondersteund door de gelegenheden- en de verplaatsingstheorie en zijn ze in lijn met eerder onderzoek. Uit een literatuurstudie die we in het kader van het onderzoek deden, bleek dat de cannabismarkt snel reageert op wijzigingen in beleid (Caulkins, 2007; Kilmer & Hoorens, 2010). Ook bleek eerder al dat drugstoeristen grotendeels wegblijven uit Nederland als ze geen toegang meer hebben tot coffeeshops, zeker als de politie intensief handhaaft op de illegale markt. (Lichte) verplaatsingseffecten zijn eveneens in ander onderzoek waargenomen (Van Ooyen e.a., 2014). Dat de coffeeshops de nieuwe criteria in principe zouden naleven, viel eveneens te verwachten op basis van eerder onderzoek (De Bruin e.a., 2007; Bieleman e.a., 2012). In onderzoek van Korf e.a. (2011) gaf een groot deel van de ingezeten coffeeshopklanten al aan dat ze zich bij invoering van een besloten-clubcriterium niet als lid van een coffeeshop zouden laten registreren.

Rest de vraag of er een derde variabele in het spel kan zijn die de veranderingen heeft bewerkstelligd. Dit lijkt niet het geval. Wel zijn er variabelen die het effect beïnvloeden. Aan de introductie van de nieuwe criteria was een communicatiecampagne gekoppeld richting niet-ingezetenen (flyers en waarschuwingborden langs de toegangswegen: ‘new rules, no drugs’). Deze kan het wegblijven van drugstoeristen bevorderd hebben. Verder was er veel media-aandacht, vooral voor het ‘oproer’ bij coffeeshopklanten en -houders in het zuiden (‘weg met de wietpas’). Met minder ophef waren de veranderingen wellicht minder groot geweest. De beschikbaarheid van extra politiecapaciteit voor de bestrijding van de illegale markt heeft invloed op de beweging van de cannabismarkt. Verplaatsing naar de illegale gebruikersmarkt wordt erdoor gehinderd en dit kan niet-ingezetenen ertoe bewegen om in eigen land te blijven.

## Conclusie

Na de implementatie van het besloten club- en het ingezetenen criterium voor coffeeshops in de drie zuidelijke provincies in mei 2012 werden in evaluatieonderzoek substantiële veranderingen waargenomen in de cannabisgebruikersmarkt en de ervaren overlast. Het drugstoerisme nam sterk af, de coffeeshops raakten zowel niet-ingezetenen als een flink deel van hun ingezeten klanten kwijt en de illegale gebruikersmarkt groeide. Omwonenden, die eerst vooral overlast ervoeren van coffeeshops, hadden meer overlast van dealactiviteiten op straat.

In het evaluatieonderzoek werd gebruik gemaakt van de fasegewijze uitrol van de nieuwe criteria, waardoor een natuurlijk experiment ontstond. Initiële verschillen tussen de experimentele groep in het zuiden en de vergelijkingsgroep in de rest van het land en de gebrekkige en uiteenlopende lokale implementatie maakten het causaal toeschrijven van de veranderingen aan de nieuwe criteria echter problematisch. In dit artikel is deze causaliteitsproblematiek onder de loep genomen.

‘The challenge for the researcher is to recognize such natural experiments when they occur and to be prepared to deal with the opportunities and challenges they present’, aldus Murnane en Willett (2011). Het natuurlijk experiment is in dit onderzoek onderkend en gebruikt, maar de uitdagingen waren complex.

De kritische beschouwing van het onderzoek in dit artikel voert ons nu tot de conclusie dat – ondanks de methodologische manco’s van dit natuurlijk experiment – zeker wel kan worden gezegd dat het besloten club- en het ingezetenen criterium de veranderingen in de cannabisgebruikersmarkt hebben *veroorzaakt*. Dit is te danken aan de brede opzet van het onderzoek, met voor- en nametingen bij de experimentele groep en een vergelijkingsgroep, multi-method-dataverzameling, *realist* inbedding en de mogelijkheid om de ontwikkelingen na de afzwakking van de interventie bij T2 te meten. Omdat de aanwezigheid van extra politiecapaciteit voor bestrijding van de illegale markt, de communicatiecampagne richting drugstoeristen en de media-aandacht voor de nieuwe criteria de veranderingen hebben beïnvloed, weten we alleen niet precies hoe groot het effect was. Met de combinatie van alle *evidence* lijkt het ons echter onwaarschijnlijk dat we met de causale conclusie een ‘type I’-fout maken.

Bij gecompliceerde natuurlijke experimenten biedt, zo is hier gebleken, een brede onderzoeksopzet kansen om tot causale conclusies te komen. Daar moet bij aangetekend worden dat de casus van de ‘wietpas’, waar het hier over gaat, wel uitzonderlijk dynamisch en controversieel was.

## Literatuur

- Bieleman, B., Nijkamp, R. & Bak, T. (2012). *Coffeeshops in Nederland 2011*. Groningen/Rotterdam: Bureau Intraval.
- Bowers, K., Johnson, S., Guerette, R., Summers, L. & Poynton, S. (2011). Spatial displacement and diffusion of benefits among geographically focused policing initiatives. *Campbell Systematic Reviews*, 3.
- Braga, A.A., Papachristos, A. & Hureau, D. (2012). Hot spots policing effects on crime. *Campbell Systematic Reviews*, 8.
- De Bruin, D., Dijkstra, R. & Breeksema, J. (2007). *Coffeeshops in Nederland 2007; naleving en handhaving van coffeeshopregels*. Utrecht: CVO.
- Caulkins, J.P. (2007). The need for a dynamic drug policy. *Addiction*, 102(1), 4-7.
- Kilmer, B. & Hoorens, S. (2010). *Understanding illicit drug markets, supply-reduction efforts, and drug-related crime in the European Union*. Santa Monica: RAND Corporation.
- Korf, D., Broekhuizen, J., Doekhie, J., Wouters, M., Boers, J. & Slot, J. (2011). *Samenvatting onderzoek coffeeshops Amsterdam*. Amsterdam: Gemeente Amsterdam.
- Van Laar, M., Ooyen-Houben, M. van & Monshouwer, K. (2009). Scheiding der markten en beleid ten aanzien van coffeeshops. In: M. van Laar & M. van Ooyen-Houben (red.). *Evaluatie van het Nederlandse drugsbeleid*. Utrecht/Den Haag: Trimbos-instituut/WODC, 109-152.
- Leeuw, F.L. (2012). Linking theory-based evaluation and contribution analysis: three problems and a few solutions. *Evaluation*, 18(3), 348-363.
- Murnane, R.J. & Willett, J.B. (2011). *Methods matter. Improving causal inference in educational and social science research*. Oxford: University Press.

Marianne van Ooyen-Houben, Bert Bieleman, Dirk Korf & Kristof De Witte

- Van Ooyen-Houben, M., Bieleman, B. & Korf, D. (2013). *Het besloten club- en het ingezetenen-criterium voor coffeeshops. Evaluatie van de implementatie en de uitkomsten in de periode mei-november 2012*. Den Haag: WODC.
- Van Ooyen-Houben, M., Bieleman, B. & Korf, D. (2014). *Coffeeshops, toeristen en lokale markt. Evaluatie van het besloten club- en het ingezetenen-criterium voor coffeeshops*. Den Haag: WODC.
- Pawson, R. (2013). *The science of evaluation: a realist manifesto*. London: Sage.
- Pawson, R. & Tilley, N. (1997). *Realistic evaluation*. London: Sage.
- Projectgroep 'Klem van gedogen' (2004). *Klem van gedogen, positief gedogen: een handelingsstrategie in de politiepraktijk*. Dordrecht: Stichting Maatschappij, Veiligheid en Politie.
- Swanborn, P.G. (2005). *Methoden van sociaal-wetenschappelijk onderzoek: nieuwe editie*. Den Haag: Boom Lemma uitgevers.
- TK 24077-259 (2011). *Drugbeleid. Brief ministers over het drugsbeleid*. Den Haag: Sdu Uitgevers.
- TK 24077-280 (2012). *Drugbeleid. Motie van het lid Kooiman c.s. over advies vraagstuk hoe het huidige drugsbeleid en de voorgenomen maatregelen zo min mogelijk schade opleveren voor de volksgezondheid*. Den Haag: Sdu Uitgevers.
- TK 24077-293 (2012). *Drugbeleid. Brief van de minister van Veiligheid en Justitie*. Den Haag: Sdu Uitgevers.
- TK 30050-2 (2005). *Handhaven en gedogen. Rapport*. Den Haag: Sdu Uitgevers.
- TK Handelingen 58-7 (2012). *Drugsbeleid. Debat over de hoofdlijnen van het drugsbeleid*. Den Haag: Sdu Uitgevers.
- TK Handelingen 69-8 (2012). *Drugsbeleid. Voortzetting van het debat over de hoofdlijnen van het drugsbeleid*. Den Haag: Sdu Uitgevers.
- Vaessen, J. (2010). *Methodological and conceptual challenges of evaluating the impact of development interventions*. Maastricht: Universitaire Pers.
- Welsh, B.C. & Farrington, D. (eds.) (2007). *Preventing crime: what works for children, offenders, victims and places*. New York: Springer.

## Bijlage

**Tabel A** *Berekening betrouwbaarheidsmarges puntschatting bij figuur 4, aantal coffeeshopbezoeken*

Lijn	Toets	Meting	M	Z	sd	n	Betrouwbaarheidsmarges		
							Onder	Gem.	Boven
	Intervalschatting; de betrouwbaarheidsmarges voor het gemiddeld aantal bezoekers is berekend met de volgende formule: $(M \pm (Z*(sd/\sqrt{n}))) * n$ .	T0	6.288	1,96	2.807	14	75.838	88.032	100.226
		T1	1.776	1,96	1.138	12	17.088	21.312	25.547
Blaauw	M staat hierbij voor de gemiddelde puntschatting per observatiegebied, Z voor de z-score bij 95% betrouwbaarheid, sd voor de gepoolde standaarddeviatie en n voor het aantal observatiegebieden.	T2	3.519	1,96	1.306	13	40.261	45.747	51.233

**Tabel B** *Berekening betrouwbaarheidsmarges puntschatting bij figuur 5, aantal coffeeshopbezoeken*

Lijn	Toets	Groep	M	Z	sd	n	Betrouwbaarheidsmarges		
							Onder	Gem.	Boven
	Intervalschatting; de betrouwbaarheidsmarges voor het gemiddeld aantal bezoekers is berekend met de volgende formule: $(M \pm (Z*(sd/\sqrt{n}))) * n$	Exp.	6.288	1,96	2.807	14	75.838	88.032	100.226
Blaauw	M staat hierbij voor de gemiddelde puntschatting per observatiegebied, Z voor de z-score bij 95% betrouwbaarheid, sd voor de gepoolde standaarddeviatie en n voor het aantal observatiegebieden.	Controle	3.211	1,96	1.845	17	44.863	54.587	64.311