

# ‘Toezicht en effectmeting. Het kán!’

*Researchsynthese van onderzoek naar  
de effectiviteit van toezichtsinterventies*

Januari 2010

Faculteit Rechtsgeleerdheid  
Vakgroep Bestuursrecht en Bestuurskunde  
Oude Kijk in 't Jatstraat 26  
9712 EK Groningen  
<http://www.rug.nl/rechten/faculteit/vakgroepen/bes>

Pro Facto BV  
Oude Boteringestraat 17a  
9712 GC Groningen  
[www.pro-facto.nl](http://www.pro-facto.nl)

© 2010 WODC, ministerie van Justitie. Auteursrechten voorbehouden.

Dit rapport is uitgebracht in opdracht van het Wetenschappelijk Onderzoek- en Documentatiecentrum (WODC) te Den Haag.

# Inhoudsopgave

Voorwoord

Hoofdstuk 1	Inleiding	1	
	1.1	Aanleiding en probleemstelling	1
	1.2	Aanpak	3
	1.3	Omschrijving toezicht	3
	1.4	Effectiviteit van toezicht	5
	1.5	Onderzoeksdiseins	7
	1.6	Screening van de studies	9
Hoofdstuk 2	Analyse	11	
	2.1	Inleiding	11
	2.2	Nalevingseffecten en maatschappelijke effecten	11
	2.3	Scherpte van doelstellingen	14
	2.4	Effecten van normstelling en effecten van toezicht	16
	2.5	Objecten van toezicht	17
	2.6	Samenstelling onderzoeksgroep en controlegroep	18
	2.7	Soorten toezichtsinterventies	20
	2.8	Verschillende methoden van dataverzameling	22
Hoofdstuk 3	Handreikingen	25	
	3.1	Inleiding	25
	3.2	Hou het eenvoudig	25
	3.3	Formuleer verwachtingen	27
	3.4	Inbedding in de toezichtsstrategie	28
	3.5	Doe ook kwalitatief onderzoek	29
	3.6	Tijdreeksmeting kan soms ook	30
	3.7	Controleer samenstelling controlegroep	31
	3.8	Let op neveneffecten	32
Bijlage 1	Samenstelling begeleidingscommissie	33	
Bijlage 2	Overzicht effectiviteitsstudies	35	



# Voorwoord

De ambitie meer te weten van de effectiviteit van toezicht en de maatschappelijke effecten daarvan is in 2008 door de Inspectieraad vertaald in een meerjarenprogramma 'Evaluatie effecten van toezicht'. Doel van dat programma is de toezichtspraktijk te verbeteren door gebruik te maken van de resultaten van onderzoek naar de effectiviteit van toezicht.

Het Wetenschappelijk Onderzoeks- en Documentatiecentrum (WODC) van het ministerie van Justitie heeft in dat kader een onderzoek uitbesteed dat als doel heeft het uitvoeren van een literatuurstudie waarin de beschikbare (wetenschappelijke) kennis in kaart wordt gebracht over de relatie tussen handhaving en toezicht, gedrag en de effecten. De literatuurstudie beoogt designs en best practices te verzamelen en te aggregeren.

In een samenwerkingsverband tussen de Vakgroep Bestuursrecht en Bestuurskunde van de Faculteit Rechtsgeleerdheid van de Rijksuniversiteit Groningen en bestuurskundig en juridisch onderzoeks- en adviesbureau Pro Facto is deze literatuurstudie uitgevoerd door Heinrich Winter en Ko de Ridder, met medewerking van Anneke de Jonge. Het onderzoek is begeleid door een commissie, waarvan de samenstelling in bijlage 1 is opgenomen. Wij bedanken de begeleidingscommissie voor de interessante gedachtewisselingen en nuttige suggesties. Een tussenresultaat van het onderzoek werd op 5 november 2009 gepresenteerd tijdens het symposium 'Handhaving en gedrag' te Utrecht.

Groningen, januari 2010

Dr. H.B. Winter  
Prof. dr. J. de Ridder



# Inleiding

### 1.1 Aanleiding en probleemstelling

Om zich te kunnen verantwoorden naar het parlement en het bredere publiek moeten toezichthouders hun prestaties inzichtelijk maken en duidelijk maken wat de meerwaarde is van hun werk. Bij effectiviteit van toezicht gaat het niet alleen om het afleggen van verantwoording. In de eerste plaats dient informatie over effectiviteit van toezicht er toe dat inspecties van hun toezichthoudende werkzaamheden leren. De Nederlandse inspecties zijn steeds meer bezig met het vergroten van het inzicht in de effecten van de eigen inspanningen. Toch is nog slechts sprake van eerste stappen. Voor zover de inspecties zelf onderzoek doen, gaat dat onderzoek vooral over de eigen prestaties en veel minder vaak over de gevolgen van de interventies van toezichthouders. Bovendien zijn de bevindingen uit dergelijk onderzoek niet op geaggregeerd niveau beschikbaar. Er zijn meerdere soorten effecten te onderscheiden. In de eerste plaats nalevingseffecten en maatschappelijke effecten. In de tweede plaats neveneffecten. De interventies van toezichthouders die tot die effecten (moeten) leiden zijn ook in verschillende categorieën te onderscheiden van communicatie en voorlichting en het verstrekken van informatie tot en met sanctionering.

De Algemene Rekenkamer deed in 2005 onderzoek naar de 'state of the art' in de wereld van het toezicht. De vraag daarbij was hoe de inspecties functioneren, hoe doelgericht de handhavingsactiviteiten zijn, én welke resultaten die inspanningen opleveren (Algemene Rekenkamer, *Handhaven en gedogen*, Tweede Kamer 2004-2005, 30 050 nr 2, Den Haag 2005). Volgens de Algemene Rekenkamer gaan de instanties tegenwoordig over het algemeen redelijk professioneel te werk. Doorgaans hebben ze wel enig inzicht in de mate van naleving op het betreffende beleidsterrein. De meeste handhavers doen aan risicoanalyse. Ze hebben alleen geen inzicht in de relatie tussen handhaving, naleving en risicoreductie, en daardoor is de fundering van hun handavingsplannen vaak gebrekkig. De inspecties maken over het algemeen bewust keuzen bij zowel handhaven als gedogen: men spitst de activiteiten steeds vaker toe op *kernbepalingen* uit de regelgeving en men oriënteert zich meer en meer op de doeleinden achter de te handhaven regelgeving. Maar de programmering van de handhaving blijft steken als het gaat om de uitdrukkelijke vaststel-

ling van het nalevingsniveau dat men wil nastreven. De hoofdconclusie van de ARK was dat er, ondanks de vele verbeteringen in de praktijk van de handhaving, nog steeds enkele belangrijke gebreken bestaan. Eén van die problemen was dat toezichthouders te weinig weten van de effecten van hun inspanningen op de naleving en op de achterliggende maatschappelijke problemen. In de evaluatie uit 2008 stelt de ARK dat nog niet bijster veel is veranderd (Algemene Rekenkamer, *Handhaven en gedogen. Terugblik 2008*, Tweede Kamer 2007-2008, 30 050 nr 5, Den Haag 2008). Wel stelt de ARK vast dat de nalevingsmetingen in de periode 2005-2007 verfijnder zijn geworden en dat er meer met de resultaten wordt gedaan.

“Maar voor een goed inzicht in alle sectoren en doelgroepen – vooral waar het hele controledomein niet jaarlijks gecontroleerd kan worden – zijn de gegevens nog niet altijd voldoende betrouwbaar.”

Over effectiviteitsonderzoek stelt de ARK dat de inspecties het meten en beoordelen van nalevingsniveaus nog vaak ten onrechte zien als een vorm van effectmeting. Ten onrechte, want wil men daadwerkelijk effecten van toezicht vaststellen dan dient ook het verband tussen (veranderingen in) nalevingsniveaus en het optreden van de toezichthouder aangetoond te worden. Effectiviteitsonderzoek is van een nóg hoger niveau als kan worden bepaald in welke mate handhaving bijdraagt aan de oplossing van *achterliggende maatschappelijke problemen*, of het voorkómen van *maatschappelijke risico's*.

De Inspectieraad vertaalde de ambitie om meer te weten te komen over de effectiviteit van toezicht in 2008 in een meerjarenprogramma 'Evaluatie effecten van toezicht'. Doel van dat programma is de toezichtspraktijk te verbeteren door gebruik te maken van de resultaten van onderzoek naar de toezichtspraktijk. In dat kader is deze literatuurstudie uitgevoerd waarin binnenlands en buitenlands effectiviteitsonderzoek is verzameld, op basis waarvan de beschikbare (wetenschappelijke) kennis over de relatie tussen toezicht en effecten in kaart is gebracht.

### **Probleemstelling**

Het bovenstaande leidt tot de volgende probleemstelling:

Welke relevante effectonderzoeken zijn uitgevoerd over toezicht en handhaving, hoe zijn die onderzoeken uitgevoerd, welke designs zijn gebruikt, wat zijn de uitkomsten en hoe zouden toezichthouders deze onderzoeken kunnen uitvoeren?

Het onderzoek leidt tot een overzicht van studies naar effectiviteit van toezicht en handhaving. Dat overzicht dient er vooral toe lessen te trekken die door Nederlandse toezichthouders kunnen worden benut voor het doen van onderzoek naar de effectiviteit van hun toezichtsuitoefening. Daarover wordt in dit verslag gerapporteerd. Behalve dat komt het overzicht ook ten goede aan de 'bibliotheek' van 'toezicht en effect': [http://www.inspectieloket.nl/vernieuwing\\_toezicht/toezichtmeteffect/bibliotheek](http://www.inspectieloket.nl/vernieuwing_toezicht/toezichtmeteffect/bibliotheek).

## 1.2

### Aanpak

De kern van de literatuurstudie bestond uit het raadplegen van tal van bronnen in het Nederlandse en het Engelse taalgebied; zowel witte als grijze literatuur. Daarbij is gebruik gemaakt van verschillende verzamelwebsites – onder andere die van de Campbell Collaboration<sup>1</sup> en van PubMed<sup>2</sup>. Vervolgens is op basis van de aldus gevonden bronnen verder gezocht met behulp van de *sneeuwbalmethode*. De database met elektronische tijdschriften van de Rijksuniversiteit Groningen bleek daarbij een onmisbaar ontsluitingsmiddel. Voorts zijn enkele sleutelfiguren uit de wereld van het toezicht benaderd en is hen gevraagd naar suggesties. Via de LinkedIn-groepen ‘toezicht met effect’ en ‘toezicht en handhaving’ (1.000+) is deelnemers verzocht suggesties van uitgevoerd onderzoek aan te leveren.

## 1.3

### Omschrijving toezicht

Toezicht omvat een brede verzameling van bestuurlijke arrangementen die gemeenschappelijk hebben dat een actor, de toezichthouder, de mogelijkheid (doorgaans de bevoegdheid) heeft om zo nodig in te grijpen in de besluiten of handelingen van andere actoren, de onder toezicht gestelden, zonder dat sprake is van een hiërarchische verhouding. Deze twee elementen (de bevoegdheid tot ingrijpen en de afwezigheid van hiërarchie) zijn bepalend voor de aard van de relatie tussen toezichthouder en onder toezicht gestelde. ‘Geen hiërarchie’ betekent dat het gaat om relaties waarin de onder toezicht gestelde zelf (primair) verantwoordelijk is en blijft voor de keuzen en handelingen die onder toezicht staan. Instructie vooraf bij voorbeeld maakt geen onderdeel uit van een toezichtsrelatie. Wel kan de toezichthouder (en dat is het eerste element) de keuzen en gedragingen van onder toezicht gestelden zo nodig beïnvloeden. Dat kan op specifieke wijze, door middel van sancties, maar ook op generieke wijze, onder meer door middel van communicatie en voorlichting. Soms kiest de toezichthouder daarbij nadrukkelijk de rol van adviseur.

In politiek en ambtelijk spraakgebruik wordt vaak de volgende begripsomschrijving gebruikt.

Toezicht is het verzamelen van de informatie over de vraag of een handeling of zaak voldoet aan de daaraan gestelde eisen, het zich daarna vormen van een oordeel daarover en het eventueel naar aanleiding daarvan interveniëren.

Met deze omschrijving worden de modules of deelprocessen aangeduid waaruit een toezichtsarrangement of een toezichtsproces is opgebouwd. Om daadwerkelijk te kunnen ingrijpen in de keuzen of handelingen van onder toezicht gestelden heeft de toezichthouder informatie nodig over die keuzen en handelingen. De vraag is hoe deze informatie wordt verworven. Een inval door de Nma bij deelnemers aan een kartel is van heel andere orde dan de vaststelling van een snelheidsovertreding met behulp van een snelheidsradar of de

---

<sup>1</sup> [www.campbellcollaboration.org](http://www.campbellcollaboration.org).

<sup>2</sup> <http://www.ncbi.nlm.nih.gov/pubmed>.

jaarlijkse opgave van de 'leggers' aan de Nederlandse Bank, waartoe verzekeringsbedrijven verplicht zijn – maar het is allemaal het verzamelen van informatie. Om zinvol te kunnen interveniëren zal de toezichthouder zich voorts een oordeel moeten vormen over de keuzen en handelingen waarover informatie is verzameld. Opnieuw is meteen de vraag hoe zo'n oordeel tot stand komt – op grond van welke maatstaven en met behulp van welke technieken. In het radarkanon zit de oordeelsvorming al ingebouwd, maar DNB moet ingewikkelde bewerkingen verrichten om op grond van de leggers te kunnen vaststellen wat de solvabiliteit van de verzekeringsmaatschappij is.

Een verwante definatorische kwestie betreft het onderscheid tussen 'toezicht' en 'handhaving'. In het spraakgebruik worden beide termen vaak door elkaar gebruikt, of men hanteert vrij willekeurige onderscheidingen. Een zinvol onderscheid vindt zijn aangrijpingspunt bij de wijze waarop het oordeel tot stand komt. De term 'handhaving' impliceert dat er iets is om te handhaven, i.c. regels. De essentie van handhaving is dan, dat een actor het vermogen (de bevoegdheid) heeft andere actoren uit een gedefinieerde verzameling (de normadressaten) ertoe te bewegen zich overeenkomstig bepaalde regels te gedragen. Bij dat vermogen (de bevoegdheid) hoort de toepassing van sancties die deze gedragsverandering beogen af te dwingen. Sancties kunnen in verschillende categorieën worden onderscheiden: er zijn bestuursrechtelijke, strafrechtelijke en civielrechtelijke sancties. De straffen kunnen ook divers zijn, variërend van financiële sancties en vrijheidsstraffen tot de toepassing van een herstelsanctie als bestuursdwang.

Uit het voorgaande blijkt dat de in bestuurlijke stukken gangbare aanduiding van toezicht geen definitie is in de zin van het afbakenen van 'toezicht' als bestuurlijk arrangement ten opzichte van andere typen bestuurlijke arrangementen. Zo zou ook het optreden van een manager beschreven kunnen worden met behulp van de drie categorieën 'informatie', 'oordeelsvorming' en 'interventie'. Niettemin is het zinvol om juist toezichtsarrangementen en toezichtsprocessen uiteen te leggen met behulp van deze drie categorieën. Elk van de drie roept een eigen set van vragen op die beantwoord moeten worden: door de ontwerper van een toezichtsarrangement en door degenen die leiding geven aan het uitvoeren van toezicht, maar ook door onderzoekers die willen vaststellen hoe toezicht in de praktijk functioneert.

Juist ook voor onderzoek naar effecten van toezicht is de genoemde driedeling van belang. Het streven van dat onderzoek is immers, om (veranderingen in) gedrag van onder toezicht gestelden in verband te brengen met het optreden van de toezichthouder. Het ligt voor de hand om gevolgen te koppelen aan interventies van de toezichthouder: deze hebben het doel gedragsverandering te bewerkstelligen. Echter, ook acties in verband met informatieverwerving en oordeelsvorming kunnen effecten bij de onder toezicht gestelden te weeg brengen. Bovendien zou de mate van effectiviteit van interventies in verband gebracht kunnen worden met de kwaliteit van de informatie waarover de toezichthouder beschikt en met de oordelen die daaraan verbonden werden.

Het belang van een consequente begripsvorming blijkt bijvoorbeeld uit de omschrijving van doeleinden van toezicht die te vinden is in de thans vigerende *Kaderstellende visie op toezicht*:

“Het is een middel om te achterhalen of burgers, bedrijven en overheden zich houden aan de gestelde regels, om misstanden te signaleren en het is een instrument om vertrouwen bij burgers, belanghebbenden en onder toezicht staande organisaties te scheppen.”

Toezichthouders zouden volgens deze aanduiding dus moeten volstaan met te registreren of de onder toezicht gestelden zich aan de norm houden. Daarmee zou de uitoefening van toezicht beperkt blijven tot de eerste twee van de drie elementen van een toezichtsproces.

## 1.4 Effectiviteit van toezicht

In algemene zin kan ‘effectiviteit van toezicht’ omschreven worden als de mate waarin de uitoefening van toezicht leidt tot verwezenlijking van de doeleinden van dat toezicht. Dit roept natuurlijk de vraag op wat dan ‘doeleinden van toezicht’ zijn.

In het voorgaande is de betekenis van ‘effectiviteit van toezicht’ impliciet al aangeduid als ‘gedragsverandering bij onder toezicht gestelden en normadressaten’. Daarbij dient aangetekend dat toezicht ook en vaak vooral de *continuering* van gedrag als doel heeft. Bij veel reguleringsregimes zijn de meeste normadressaten geneigd tot normnaleving. Een geringe druk van de toezichthouder kan dan al volstaan om de normnaleving voort te zetten. Veel onderzoek beoogt de werking vast te stellen van een bepaald toezichtsoptreden op een bepaald moment in termen van een bepaalde *gedragsverandering*. De effectiviteit van toezicht in termen van het *onderhouden* van bepaalde gedragspatronen onder de normadressaten is minder vaak onderzocht. Longitudinaal onderzoek is daarvoor noodzakelijk, maar niet voldoende. Ook bij metingen in een tijdreeks komt immers nog niet onbetwist vast te staan dat de gemeten verandering aan het toezicht moet worden toegeschreven.

Een tweede aantekening betreft de aard van de gedragingen die de toezichthouder beoogt te bevorderen. Zoals hiervoor al aangegeven is de aard van dat gedrag vaak vastgelegd in regels en normen. De toezichthouder heeft dan als doel om *naleving* van de regels en normen te bewerkstelligen. Er zijn ook situaties waarin een toezichthouder geen regels of normen tot zijn beschikking heeft. Hij zal dan de (on)aanvaardbaarheid van waargenomen gedragingen of beslissingen moeten afleiden uit de verder weg gelegen doeleinden van het toezicht: de maatschappelijke doeleinden. Daarnaast zijn er situaties waarin de toezichthouder weliswaar over regels beschikt, maar waarin naleving van die regels niet zijn eerste doel is. De toezichthouder oriënteert zich dan in de eerste plaats op het bereiken van de maatschappelijke doelen. Omgekeerd richten andere toezichthouders zich in de eerste plaats op naleving van regels en normen en is de (vaak plausibele) veronderstelling dat daarmee uiteindelijk ook de maatschappelijke doelstellingen gehaald worden. Als het gaat om effecten van het optreden van toezichthouders kan dus worden onderscheiden tussen *nalevingseffecten* en *maatschappelijke effecten*. Toezichthouders kunnen zich richten op het *versterken van naleving*, maar ook op het behalen van *maatschappelijke doelen*.

Onderzoek onder milieuhandhavers liet zien dat zij er vooral op zijn gericht om het maatschappelijke probleem (milieuschade) te voorkomen of te verminderen. Darto spreken zij degenen die milieuschade veroorzaken of zouden kunnen veroorzaken aan op hun gedrag. De wettelijke regels die door dat gedrag worden geschonden gebruiken zij daarbij dan als hulpmiddel.<sup>3</sup>

De relatie tussen regelgeving enerzijds en toezicht en handhaving anderzijds heeft nog een andere problematische kant. Regels zijn soms en tot op zekere hoogte *self-executing*: het gedrag van de normadressaten wordt rechtstreeks beïnvloed door normering zelf. In hoeverre zulks het geval is, is een onderwerp van onderzoek en debat in de rechtssociologie. Maar dat de effecten van toezicht vaak worden versluierd door effecten van regelgeving lijkt geen onjuiste veronderstelling. Methodisch gezien is de werking van de regelgeving niet anders dan de werking van elke andere interveniërende variabele. Het onderzoeksdesign zal er in moeten voorzien voor deze variabele te controleren.

Een derde kanttekening betreft de ‘maatschappelijke doelen’ van toezicht. Hiervoor werd al het onderscheid tussen verschillende soorten effecten genoemd. We maakte een onderscheid tussen *nalevingseffecten* en *maatschappelijke effecten* en wezen ook op neveneffecten. Bij maatschappelijke effecten gaat het om de verwezenlijking van maatschappelijke doelen die het veronderstelde gevolg zijn van de naleving van gestelde regels. In de praktijk is er vaak een veelheid en een gelaagdheid van dat soort doeleinden te bespeuren, zodanig dat het geheel van maatschappelijke doeleinden is af te beelden in een doelboom.

Zo is het toezicht op overnamebiedingen van de AFM en zijn Europese collega’s er op gericht dat bedrijven bij overnames de in de Europese richtlijn inzake het openbaar bod vervatte regels naleven. Het einddoel van deze regulering is het in stand houden van een goed functionerende markt voor aandelen. Tussendoelen zijn onder meer: het bewerkstellingen van een ‘level playing field’ voor kopers en verkopers en gelijke behandeling van alle aandeelhouders in verhandelde ondernemingen. Om dat te bereiken dienen onder meer potentiële kopers transparant te zijn in hun biedingen en moet de relevante informatie voor alle belanghebbenden tijdig en gelijktijdig aanwezig zijn. Vanuit het perspectief van de Europese Unie is er daarenboven nog een verder gelegen doel, namelijk dat de regulering inclusief de handhaving, in de gehele Unie op dezelfde wijze gebeurt om te bewerkstellingen dat er één interne markt zonder hinderpalen tot stand komt of in stand blijft.<sup>4</sup>

Een laatste kanttekening betreft het verschil tussen ‘effecten’ en ‘effectiviteit’ en daarbij komen de eerder genoemde neveneffecten aan de orde. Het optreden van de toezichthouder kan allerlei gevolgen hebben, die daarom als effecten zijn aan te merken. Omdat het begrip ‘effectiviteit’ een koppeling legt met de doelstellingen van het optreden, verwijst dat begrip enkel naar die effecten die *beoogd* zijn. Niet-beoogde effecten, *neveneffecten*, kunnen zowel negatief als positief zijn en zelfs de effectiviteit van het toezicht on-

---

<sup>3</sup> J. de Ridder, N. Struiksma: *De kern van de zaak*. Groningen 2008.

<sup>4</sup> A. Haan-Kamminga, *Supervision on takeover bids, a comparison of regulatory arrangements*. Deventer, Kluwer: 2006.

dermijnen. Een voor toezicht typerend negatief neveneffect is 'moral hazard': indien men eventuele negatieve consequenties van eigen gedrag kan afwentelen op anderen (bij voorbeeld een verzekeringsmaatschappij) of de samenleving als geheel, zal men grotere risico's nemen dan indien men eventuele schade zelf moet dragen. Op gelijksoortige wijze kan de perceptie, dat de toezichthouder waakt ertoe leiden dat de samenleving, althans belanghebbenden, zelf minder goed risico's afwegen en daardoor minder verantwoorde keuzen doen. Als gevolg van het optreden van de toezichthouder, (of rechtstreeks als gevolg van de normering in de wetgeving), nemen onder toezicht gestelden de beoogde maatregelen, maar tegelijkertijd veranderen ze hun gedrag waardoor het effect van die maatregelen wordt ondermijnd. Dat verschijnsel komen we bijvoorbeeld tegen in veel onderzoek naar maatregelen die beogen de verkeersveiligheid te bevorderen. Sommige verkeersdeelnemers gaan veiligheidsgordels dragen, maar vertonen tegelijkertijd roekelozener rijgedrag.

Met deze kanttekeningen is de reikwijdte van wat onder 'effecten van toezicht' verstaan kan worden wel aangegeven. Een vraag voor de analyse is welke keuzen dienaangaande in de te bestuderen onderzoeken zijn gemaakt.

## 1.5 Onderzoeksdesigns

De vaststelling van effecten van toezicht, hoe ook verder gespecificeerd, vraagt zoals gezegd in elk geval om het vaststellen van een verband tussen gedragingen van onder toezicht gestelden en het optreden van de toezichthouder. Wanneer een causale relatie kan worden gevonden tussen de handelingen van de toezichthouder en het gedrag of de gedragsveranderingen van de onder toezicht gestelden, spreken we van toezichtseffecten. Effectstudies zullen over het algemeen niet alle effecten in kaart willen brengen maar zich vaak richten op enkele mogelijke gevolgen van de uitoefening van toezicht.

Van die mogelijke gevolgen zullen de *beoogde* effecten doorgaans het meeste aandacht krijgen. Onderzoek gericht op de vraag of de beoogde gevolgen worden gerealiseerd heet effectiviteitsonderzoek. Effectiviteitsonderzoek combineert in de meest strikte benadering vier soorten waarnemingen. In de eerste plaats waarnemingen vóór de interventie en waarnemingen ná de interventie. De vergelijking van beide kan laten zien of er een gedragsverandering heeft plaats gevonden na de toezichtsuitoefening. Een alternatief waarvoor – bij het ontbreken van een waarneming voorafgaand aan de interventie – vaak wordt gekozen is het doen van waarnemingen in tijdreeksen ná de interventie. De verandering in het gedrag wordt dan gerelateerd aan het optreden van de toezichthouder. Bij een voor/na design of een design met metingen op verschillende momenten staat nog niet vast dat de waargenomen gedragsverandering het *gevolg* is van het bestudeerde toezichtsoptreden. Er zijn vaak allerlei andere causale relaties die de gedragsverandering veroorzaken. Om vast te stellen welke veranderingen aan het optreden van de toezichthouder moeten worden toegeschreven – en dus aangemerkt moeten worden als toezichtseffecten – is een tweede set waarnemingen nodig bij een groep onder toezicht gestelden (normadressaten) bij wie het te onderzoeken optreden van de toezichthouder niet plaats vindt. Ook bij deze controlegroep moet de vergelijking tussen een voormeting en een nameting of een meting in een tijdreeks uitwijzen welke veranderingen in het ge-

drag zich hebben voorgedaan. Kan bij de controlegroep geen verandering worden vastgesteld dan is de conclusie dat het toezicht het veronderstelde effect had op zijn minst aannemelijk.

Effectiviteitsonderzoek combineert dus een voor/na meting met een met/zonder benadering, waarbij de metingen plaats vinden zowel bij de onderzoeksgroep (of experimentgroep), als bij de controlegroep.

Dit soort effectiviteitsonderzoek eist onderzoeksdesigns met een hoge interne validiteit, dat wil zeggen designs waarbij de conclusie over de werking van het toezicht zeer aannemelijk is. In de literatuur wordt de Maryland Scientific Methods Scale (MSMS) gebruikt om verschillende onderzoeksdesigns te ordenen. Deze vijfpuntsschaal begint op het laagste niveau met onderzoek dat een samenhang vaststelt tussen een interventie en de gedragsverandering en eindigt bij onderzoek met een voor- en een nameting en een controlegroep, waarbij sprake is van een willekeurige samenstelling van controle- en onderzoeksgroep. De willekeurige samenstelling voorkomt een bias in de selectie van te onderzoeken normadressaten en daarmee de introductie van niet gemeten factoren – bijvoorbeeld een grotere ontvankelijkheid voor toezichtsinterventies in de onderzoeksgroep.

Tabel 1: Vijf niveaus van interne validiteit conform de Maryland Scientific Methods Scale

Niveau 1	Samenhang tussen een beleidsinterventie en uitkomstvariabele na invoering van de beleidsinterventie
Niveau 2	Score op de uitkomstvariabele voor en na invoering van de beleidsinterventie, zonder (vergelijkbare) controlegroep
Niveau 3	Score op de uitkomstvariabele voor en na invoering van de beleidsinterventie, gemeten in een experimentele en vergelijkbare controlegroep (quasi-experimenteel design)
Niveau 4	Score op de uitkomstvariabele voor en na invoering van de beleidsinterventie gemeten in een quasi-experimentele setting, met controle voor andere variabelen die de uitkomstvariabele beïnvloeden
Niveau 5	Score op de uitkomstvariabele voor en na invoering van de beleidsinterventie, waarbij de beleidsmaatregel 'at random' is toebedeeld aan een experimentele en een controlegroep (experimenteel design).

Voor dit literatuuronderzoek is gezocht naar studies naar de effectiviteit van het optreden van toezichthouders die op ten minste MSMS-3 niveau zijn uitgevoerd. Meestal gaat het bij die studies om kwantitatief onderzoek, soms gecombineerd met kwalitatieve methoden van gegevensverzameling. De reden hiervoor is uiteraard dat bij een klein aantal waarnemingen de significantie van een bepaald verband niet kan worden vastgesteld, dat verband kan immers volgens het toeval tot stand zijn gekomen. Daarnaast kunnen 'kleinere' effecten beter in kwantitatief onderzoek worden opgespoord, als het aantal waarnemingen maar voldoende groot is. Dat neemt niet weg dat kwalitatief effectiviteitsonderzoek – in ieder geval als aanvulling op kwantitatief onderzoek – ook meerwaarde kan hebben. Een veel gevolgde benadering is die van het *multiple casestudy-design*. Deze be-

nadering is bijvoorbeeld toegepast in het dissertatieonderzoek van Ehren naar de effecten van (onderwijs)toezicht.<sup>5</sup>

## 1.6 Screening van de studies

De studies zijn volgens een vast stramien geanalyseerd. Beschreven zijn het onderwerp van het onderzoek (titel van de publicatie), degene die het onderzoek uitvoerde (auteurs, instelling), het doel van het onderzoek (de probleemstelling), of het onderzoek zich richt op maatschappelijke effecten of nalevingseffecten, de opzet van het onderzoek (welk MSMS-niveau, wanneer uitgevoerd, welke onder toezicht gestelden, hoe gemeten, samenstelling experimentgroep en controlegroep), de ingezette interventies en de daarbij gebruikte toezichtsmiddelen en tot slot de resultaten van het onderzoek.

In bijlage 2 bij deze rapportage is een overzicht opgenomen van de studies die zijn aangetroffen. Niet alle studies in het overzicht voldoen helemaal aan de daarvoor gestelde eisen. Niet alle onderzoeksdesign halen het gewenste MSMS-3 niveau. In een aantal gevallen gaat het om reviews van studies waarbij niet steeds duidelijk was welk design in de bestudeerde onderzoeken was toegepast. Andere studies zijn, hoewel ze het MSMS-3 niveau niet halen, om andere redenen wel interessant. En er is ook een aantal studies van toezichtsinterventies opgenomen dat op MSMS-4 en MSMS-5 niveau zit.

Het is van belang op te merken dat voortgezet onderzoek naar studies over effectiviteit van toezicht tot een uitgebreider overzicht zal leiden. Het overzicht is dus niet uitputtend. Wel is getracht daarin in ieder geval onderzoek op te nemen dat ziet op de belangrijkste beleidsterreinen en toezichthouders, zodat lessen kunnen worden geleerd voor de Nederlandse toezichtspraktijk.

In totaal zijn 70 studies beschreven. Die zijn in een aantal beleidsterrein ingedeeld:

- AO (Arbeidsomstandigheden): 5
- B (Belastingen): 3
- EF (Economie & Financiën): 10
- O (Onderwijs): 4
- OOV (Openbare Orde en Veiligheid): 8
- M (Milieu): 4
- V (Verkeer & Vervoer): 21
- VG (Volksgezondheid): 7
- VV (Voedselveiligheid): 7

In hoofdstuk 2 wordt op geaggregeerd niveau gerapporteerd over de gevonden studies. De op basis daarvan te formuleren 'handreikingen' voor de toezichtspraktijk zijn te vinden in hoofdstuk 3.

---

<sup>5</sup> M.C.M. Ehren, *Toezicht en schoolverbetering*. Delft, Eburon: 2006.



# Analyse

### 2.1 Inleiding

In dit hoofdstuk worden de gevonden studies op een geaggregeerd niveau besproken. Weergegeven wordt in welke mate onderzoekers er in slagen onderzoek te doen naar de effectiviteit van toezicht en welke problemen ze daarbij tegenkomen. Op basis daarvan worden adviezen geformuleerd, gericht tot Nederlandse toezichthouders met het oog op het uitvoeren van onderzoek naar de effectiviteit van toezicht(sinterventies).

### 2.2 Nalevingseffecten en maatschappelijke effecten

Zoals hiervoor uiteengezet kan globaal worden onderscheiden tussen onderzoek naar twee soorten effecten van toezicht: *nalevingseffecten* en *maatschappelijke effecten*. Behalve op (verbetering van de) normnaleving is het werk van toezichthouders dikwijls ook gericht op het realiseren van de maatschappelijke effecten van de norm in kwestie. In de verzamelde studies komen beide effecten van toezicht voor als afhankelijke variabele van het betreffende onderzoek.

Nalevingstoezicht is het toezicht op handelingen van burgers, bedrijven en instellingen, waarbij het toezicht richt zich op de naleving van regelgeving. Maatschappelijke effecten zijn een stuk moeilijker te meten en gaan een stap verder dan de nalevingseffecten. Vaak zijn effectiviteitsdesigns beperkt tot onderzoek naar normnaleving. Vastgesteld wordt niet of door het optreden van de toezichthouder de uiteindelijk gewenste doelen van de regelgever worden gehaald, maar simpelweg of de onder toezicht gestelden zich conformeren aan de gestelde norm. Dat lijkt een naïeve keuze te zijn, maar die keuze heeft dikwijls te maken met het feit dat het niet altijd goed mogelijk is de maatschappelijke effecten te onderzoeken. Wellicht nog vaker is dergelijk onderzoek op zichzelf wel mogelijk, maar is het zeer tijdrovend en/of kostbaar. En niet uit te sluiten is dat toezichthouders soms eigenlijk niet precies weten welke maatschappelijke effecten met hun interventies worden beoogd.

Niet onbelangrijk is ten slotte, dat er in veel gevallen geen ander doel is dan het bevorderen van de *naleving*. Het maatschappelijke effect en het nalevingseffect vallen dan als het ware samen. Een voorbeeld daarvan is het toezicht op de naleving van de belastingwetgeving. Het doel daarvan is het bevorderen van normconformiteit. Het beoogde maatschappelijke effect daarvan is het vergroten van de belastinginkomsten, dat automatisch wordt gerealiseerd bij een grotere normconformiteit.

Een redelijk groot aantal van de onderzochte studies gaat over het toezicht op de wetgeving over verkeersveiligheid. De onderzochte maatregel is in een aantal studies het handhaven van de verplichting tot het dragen van veiligheidsgordels. Sommige studies beperken zich daarbij tot onderzoek naar de *nalevingseffecten*. Gekozen wordt dan voor een design waarbij door middel van observaties (bijvoorbeeld op kruispunten) wordt vastgesteld of verkeersdeelnemers autogordels dragen. In andere studies wordt ook nog getracht een causale relatie te leggen met het beoogde *maatschappelijke effect*, het voorkomen van verkeersslachtoffers. Ongevallencijfers van de politie vormen daarvoor veelal de basis.

Toezichthoudende instanties hebben in veel gevallen een voorkeur voor het in kaart brengen van de maatschappelijke effecten van hun werk. Daarmee kan de organisatie tegenover het politieke gezag en tegenover de samenleving als geheel het belang van het toezichtswerk laten zien. Maar vaak is het doen van onderzoek naar maatschappelijke effecten onmogelijk of wegen de extra kosten en de daarvoor benodigde tijd niet op tegen die meeropbrengst. Dat het in effectiviteitsonderzoek vaak gaat om onderzoek naar het behalen van nalevingseffecten blijkt uit de door ons verzamelde studies. In de helft van de onderzoeken (35 studies) worden uitsluitend uitspraken gedaan over de nalevingseffecten. De andere helft (34 studies) worden (ook) de maatschappelijke effecten van het toezicht in kaart gebracht. In de helft daarvan (een vierde van het totaal, 18 studies) gaat het uitsluitend om het meten van een maatschappelijk effect. De andere helft (iets minder dan een vierde, 16 onderzoeken) combineert beide benaderingen. De meeste onderzoeken concentreren zich dus op nalevingseffecten.

Een voorbeeld van een studie waarin zowel de nalevingseffecten als de maatschappelijke effecten in kaart zijn gebracht is het onderzoek, dat de Amerikaanse Federal Motor Carrier Safety Administration (FMCSA) heeft laten uitvoeren door het Volpe Center ('Saved Lives', onderzoek 28). Het maatschappelijke doel dat de *agency* zich gesteld had was het 'large-truck-and-busfatality-rate' in 2011 terug te brengen tot 1,6 slachtoffers per 100 miljoen voertuigmijlen. Dit reductiedoel past bij de politieke doelstelling de algemene verkeersveiligheid met 40 procent te verhogen ten opzichte van het basisjaar 1996 (zie ook de website van dit *agency* en het 'mission statement' van de organisatie). Het aantal serieuze incidenten met vervoer van gevaarlijke stoffen per vrachtwagen diende met 20 procent worden teruggebracht: van 574 in 2000 tot circa 460 in 2010. In deze studie is gekeken naar de effectiviteit van het optreden van toezichthouders op de verkeersveiligheid, in termen van de gestelde doelen. Uiteindelijk slaagt FMCSA er in een kosten/baten berekening te maken van de uitgevoerde inspectiewerkzaamheden waarin het aantal 'saved lives' en het verminderde aantal verkeersgewonden kan worden afgezet tegen de kosten van het toezicht.

Een voorbeeld van een studie waarin niet het effect van toezicht op de verkeersveiligheid is onderzocht maar het effect op de normnaleving, is een Amerikaanse studie in twee testgemeenten en één controlegemeente naar het gebruik van veiligheidsgordels en –zitjes voor peuters, begin jaren negentig (studie 27). Het beoogde en onderzochte effect in dat onderzoek was de mate van naleving van regels over het gebruik van veiligheidsgordels en -zitjes. In een dergelijk onderzoek wordt veelal aangenomen dat betere normnaleving leidt tot de gewenste maatschappelijke effecten. Om bij het voorbeeld te blijven: we weten uit allerlei onderzoek (bij voorbeeld botsproeven) dat wanneer kinderen niet op de voorgeschreven wijze bij vervoer in auto's zijn beveiligd, ze een grotere kans lopen op ingrijpend letsel bij een ongeval.

Een ander onderzoek waarin geen relatie met de maatschappelijke impact is gelegd, maar waarin de naleving centraal stond is het onderzoek van de Nederlandse Voedsel- en Warenautoriteit naar de naleving van hygiënevoorschriften in Chinese (afhaal)restaurants (onderzoek 38). Het onderzoek toont aan dat de naleving van deze voorschriften in de onderzoeksgroep, waar handhaving plaats vond, gecombineerd met een uitgebreid communicatietraject, significant is toegenomen in vergelijking met de naleving in de controlegroep. De onderzoekers rapporteren niet over de maatschappelijke effecten. In buitenslands onderzoek naar de relatie tussen hygiëne-inspecties en voedselveiligheid in restaurants is het soms wel gelukt die relatie met maatschappelijke effecten te leggen. In een Amerikaans onderzoek uit 1999 is getracht te voorspellen welke restaurants een grotere kans op voedselonveilige situaties hebben (studie 42). De ontbrekende stap tussen de nalevingseffecten en de maatschappelijke effecten is daar gezet door meldingen over voedselvergiftiging te gebruiken. Ook in een ander Amerikaans onderzoek, naar voedselveiligheid in Seattle-King county (studie 40), is het gelukt dat verband te leggen.

Wanneer we de totale dataset bezien, valt op dat – zoals te verwachten – onderzoek naar maatschappelijke effecten vooral is gedaan op terreinen waar die maatschappelijke impact relatief eenvoudig meetbaar is. Op het terrein van milieukwaliteit en bescherming van de visstand (studies 10, 11, 12) is het niet zo eenvoudig de maatschappelijke effecten van toezichtsinterventies in kaart te brengen. Toch is er ook een studie beschreven waarin wel is getracht een verband te leggen tussen een milieubeschermdende maatregel, in casu een verbod zich in een bepaalde strook water voor de kust op te houden, en de ontwikkeling van schelpdieren in dat gebied. In studie 9 is zelfs een vergelijking gemaakt met een controlegebied. Er is overigens wel kritiek mogelijk op het design van dat onderzoek. In de eerste plaats is het de vraag of het de onderzoekers gelukt is een controlegebied te vinden dat goed vergelijkbaar is met de kenmerken van het experimentgebied. Verder is het in samenhang daarmee door het ontbreken van een voormeting of metingen op verschillende tijdstippen na de start van het experiment erg onzeker of de gevonden verschillen in aantal en omvang van de schelpen in het experimentgebied en het controlegebied aan de interventie toegeschreven kunnen worden.

Zoals gezegd toont het overzicht aan dat onderzoekers op het terrein van het toezicht op verkeersmaatregelen er regelmatig wel in slagen een verband te leggen tussen de toezichtsinterventie en de daarmee beoogde maatschappelijke effecten. Er zijn op het terrein van verkeersmaatregelen relatief veel studies gedaan naar het effect van toezicht op de

verkeersveiligheid. De studies 13 tot en met 33 gaan over Verkeer & Vervoer. Van die 21 studies zijn er 8 onderzoeken alleen gericht op normnaleving. Onderzoekers van het effect van verkeersveiligheidstoezicht slagen er daarmee beter dan onderzoekers van andere toezichtssituaties in een relatie te leggen tussen het toezicht en de maatschappelijke impact van de normstelling, het bevorderen van verkeersveiligheid. Dat kan gemakkelijk begrepen worden. In de bedoelde studies wordt als maatstaf voor verkeersveiligheid het aantal verkeersslachtoffers gehanteerd. Daarbij gaat het om cijfers die redelijk goed worden geregistreerd en die uit bestaande databestanden kunnen worden gehaald.

Ook op het terrein van het toezicht op voedselveiligheid is het relatief eenvoudig na te gaan of inspecties van restaurants en andere instellingen waar voedsel wordt bereid leiden tot grotere voedselveiligheid. Dat kan door als variabele voor maatschappelijke effecten het aantal rapportages over incidenten rond voedselvergiftigingen te nemen. In de beide eerder genoemde Amerikaanse studies naar de maatschappelijke effecten van toezicht op voedselveiligheid (studies 40 en 42) is het de onderzoekers gelukt een verband te leggen tussen inspecties bij restaurants en het aantal gerapporteerde gevallen van voedselvergiftiging.

In de tien gerapporteerde onderzoeken over de effectiviteit van toezichtsinterventies op het terrein van Economie & Financiën lukt het, anders dan de hiervoor genoemde studies op het terrein van Verkeer & Vervoer en Voedselveiligheid, niet om een relatie te leggen met de maatschappelijke effecten. Dat is goed te begrijpen omdat effecten op de doelstellingen zoals 'transparante aandelenhandel' of 'level-playing field' veel lastiger meetbaar zijn. De onderzoeken die zich richtten op de effecten van wetgeving over het verbod op handel met voorwetenschap trachten door middel van een econometrische analyse bij benadering in kaart te brengen welke gevolgen introductie van dergelijke normen heeft voor de ontwikkeling van handelsvolumes en van aandelenprijzen in relatie tot de bekendmaking van beursgevoelige informatie. Wat dan wordt gemeten is de mate waarin de normnaleving verbetert of verandert. Of de transparantie van de aandelenmarkt groter is geworden en of er sprake is van 'eerlijker' prijsvorming is dan nog niet vastgesteld.

## 2.3 Scherpheid van doelstellingen

Om uitspraken te kunnen doen over de effectiviteit van toezichtsinterventies zijn onderzoekers in principe gebonden aan de wijze waarop de toezichthouder de doelstelling van het toezicht formuleert. Die doelformulering kan gelijk zijn aan de maatschappelijke doelstelling of de nalevingsdoelstelling die de wetgever aan zijn normering ten grondslag legt. De toezichthouder kan ook afgeleide, of geoperationaliseerde doelstellingen formuleren. Doelstellingen van toezichthouders worden niet vaak in precieze, kwantitatieve zin geformuleerd. Veelal gebeurt dat in kwalitatieve zin: 'grotere verkeersveiligheid', 'minder verkeersdoden', 'afname van de onveiligheidsgevoelens', 'betere schoolprestaties' of – zoals hiervoor besproken – 'transparante aandelenhandel' of 'level-playing field'. Wanneer doelstellingen van toezichtsinterventies wel in meer precieze, kwantitatieve zin worden geformuleerd, kan een meting ook meer precies rapporteren over de effectiviteit van de interventie in termen van het behaalde succes. Uiteraard is het zo dat wanneer de doelstellingen door de regelgever of de toezichthouder vaag zijn geformuleerd dat niet

betekent dat over het effect van de interventie geen precieze uitspraken zijn te doen. Wanneer de doelstelling 'verbetering van de verkeersveiligheid' is, kan effectiviteitsonderzoek er desondanks in slagen exact vast te stellen hoe groot de reductie van het aantal verkeersslachtoffers is als gevolg van een bepaalde toezichtsinterventie. In die situatie kunnen alleen geen uitspraken worden gedaan over de mate waarin de doelstelling is gerealiseerd.

In veel van de gevonden toezichtsonderzoeken is sprake van kwalitatieve doelstellingen: meer schelpdieren in een gebied (studie 9), betere Cito-scores van basisschoolleerlingen (studie 34) en vermindering van de onveiligheidsgevoelens (studie 3). Het hiervoor reeds aangehaalde Amerikaanse onderzoek naar het effect van bedrijfscontroles bij vrachtvervoerders op het aantal verkeersslachtoffers (studie 28) daarentegen werkte met een zodanig scherpe kwantificering van de vermindering van het aantal verkeersslachtoffers (ook tot uitdrukking komend in de titel, 'saved lives'), dat het betreffende agency in staat werd gesteld het toezichtsresultaat uiteindelijk in termen van kosten en baten te presenteren.

Globaal geformuleerde doelstellingen kunnen in evaluatieonderzoek aangescherpt worden tot nauwkeuriger meetbare effecten. Onderzoek 1, waarin de effectiviteit van cameratoezicht in het Verenigd Koninkrijk werd gemeten in 13 gebieden, gebruikt de verandering in absolute aantallen geregistreerde misdrijven als operationalisering van het beoogde effect: daling van criminaliteit in door camera's bewaakte openbare ruimten. Gelijksortige effectiviteitsmaatstaven werden gebruikt in de Amsterdamse evaluaties van cameratoezicht (onderzoeken 2 en 3). Het onderzoek naar de stand van de Ezelsoor (studie 9) gebruikte niet alleen het aantal schelpdieren maar ook de grootte van de dieren als indicator voor het effect. Bij de evaluatie van de handhaving van de maximum snelheid op een aantal wegen in de provincie Fryslân (onderzoek 13) is de doelstelling 'gedragsverandering' geoperationaliseerd in: 'veranderingen in de gemeten gemiddelde snelheid op de onderzochte wegen' en 'veranderingen in aantal bekeuringen wegens snelheidsovertredingen'. De doelstelling 'verbetering veiligheid op de weg' werd gemeten in termen van de verandering van het aantal (dodelijke) verkeersslachtoffers.

Onderzoekers kunnen dus proberen vaag geformuleerde doelstellingen van de toezichthouder of wetgever te operationaliseren in meetbare doelen om zo in staat te zijn uitspraken te doen over de effectiviteit van het toezicht. Dat houdt wel een risico in, namelijk dat de toezichthouder zich niet herkent in die operationalisering en de uitspraken vervolgens naast zich neerlegt. Het heeft daarom de voorkeur dat de toezichthouder zelf scherpe doelstellingen formuleert die aan onderzoek onderworpen kunnen worden. In een minder optimale benadering legt de toezichthouder zich voordat het onderzoek van start gaat vast op de operationalisatie van zijn vage doelstellingen door de onderzoekers.

Een ander voorbeeld van een onderzoek waarin geen relatie is gelegd met een in kwantitatieve en precieze termen geformuleerde doelstelling is studie 45. Dat onderzoek heeft als doel uitsluitend te geven over twee rivaliserende theorieën over de effecten van 'naming and shaming', in dit geval de bekendmaking van vertrouwelijke bankgegevens. Doordat tevoren twee hypothesen werden geformuleerd die in het onderzoek konden worden getoetst, is het toch mogelijk gebleken een concrete uitspraak te doen aan de

hand van de onderzoeksgegevens over de effectiviteit van de toezichtsinterventie. De hypothesen luiden in vragende vorm, kortgezegd, “werkt ‘naming and shaming’ disciplinerend op rekeninghouders of heeft ‘naming and shaming’ juist tot gevolg dat rekeninghouders ongedisciplineerd weglopen (‘bank run’)?” De onderzoeksresultaten lieten zien dat rekeninghouders in het geheel niet reageerden op deze actie van de toezichthouder.

Uit het voorgaande volgt dat doelstellingen van toezichthouders niet altijd in kwantificeerbare, meetbare termen zijn gesteld. Voor onderzoek naar de effectiviteit van toezichtsinterventies is het wel nodig dat tevoren een scherp omlinjende hypothese wordt geformuleerd, wil het onderzoek uitspraken kunnen opleveren over het succes of falen van de bestudeerde toezichtsinterventie. De bestudeerde effectiviteitsonderzoeken slagen er soms in een ‘vaag’ geformuleerde doelstelling van de toezichthouder in meetbare, geoperationaliseerde termen te vatten. Bij het opzetten van het design van het onderzoek vraagt dit onderwerp veel aandacht.

## 2.4 Effecten van normstelling en effecten van toezicht

Onderzoekers die de effectiviteit van toezichtsinterventies willen vaststellen moeten een keuze maken tussen het bestuderen van nalevingseffecten en/of maatschappelijke effecten. Daarnaast moeten ze voorafgaand aan het uitvoeren van het onderzoek bij het ontwikkelen van het onderzoeksdesign de doelstelling van de toezichtsinterventie (her)formuleren in een meetbare, liefst kwantitatieve maatstaf. Naast deze kwesties krijgen ze te maken met een causaliteitsvraag. Bij empirisch onderzoek naar het gedrag van onder toezicht gestelden is het de vraag of hun gedragsverandering bewerkstelligd wordt door de toezichtsinterventies of door de normstelling zelf dan wel door een combinatie van beide.

Als in een onderzoek een gedragsverandering wordt gevonden moet vervolgens worden achterhaald waardoor die gedragsverandering is veroorzaakt. De vraag is dus of de gevonden effecten zijn toe te schrijven aan regulering of aan de activiteiten van de toezichthouder. De onderzoeken die in het kader van deze studie zijn geanalyseerd liggen meestal ten minste op MSMS-3 niveau. Dat design voorziet in een controlegroep en dus in waarnemingen van situaties met en zonder de interventies van toezichthouders. Aangenomen kan worden dat gecontroleerd is voor de rechtstreekse werking van de normstelling. Dat betekent dat dit design in staat is de causale relatie te leggen tussen het optreden door de toezichthouder en de gevonden gedragsverandering.

In de verzameling gevonden studies is ook een aantal studies opgenomen die strikt genomen niet gaan over de effectiviteit van toezichtsinterventies, maar die de effecten van normstelling in kaart brengen. Dat geldt voor de meeste onderzoeken op het terrein van Economie & Financiën. Deze studies gaan overwegend over regelgeving over handel in aandelen. In studies 48 tot en met 52 staat de introductie van ‘insider trading’ centraal. Het gaat in die normstelling over handel met voorkennis en het bevorderen van ‘market cleanliness’. Hoewel deze studies niet gaan over het werk van toezichthouders, leveren ze wel interessante informatie op die van invloed zou kunnen zijn op de keuzes die ‘marktmeesters’ maken bij het stellen van prioriteiten. Deze onderzoeken hebben gemeen dat

ze een verband leggen met de publicatie van koersgevoelige informatie door beursgenoteerde bedrijven enerzijds en de handel in het betreffende aandeel anderzijds. Het gaat er in dat onderzoek om te achterhalen of er voorafgaand aan de publicatie van die beursgevoelige informatie (zoals bij voorbeeld kwartaalcijfers, maar het kan ook om andere bedrijfsberichten gaan) afwijkingen te zien zijn in de handel in het aandeel waarop die informatie betrekking had. Soms is ook een relatie gelegd met veranderingen in de prijs van het aandeel. Bij dat onderzoek is niet gekeken naar acties van toezichthouders, maar is een vergelijking gemaakt tussen de koersbewegingen rond dergelijke publicaties voorafgaand en na de introductie van beperkende wetgeving. Dergelijk onderzoek is van belang voor toezichthouders omdat het aanwijzingen oplevert over de kenmerken van bedrijven (groot/klein, hoog/laagtechnologisch etc) die zich meer of minder aan de normen conformeren.

Amerikaans onderzoek naar de resultaten van een nieuwe manier van toezicht op het terrein van voedselveiligheid in Los Angeles County (studie 39) wordt gepresenteerd als een effectiviteitsstudie. In het onderzoek is echter slechts een voor/na vergelijking uitgevoerd; het is dus geen MSMS-3 onderzoek. Hoewel de gemeten verandering wordt toegeschreven aan de gewijzigde manier van inspecteren, is dat geenszins zeker. Denkbaar is dat in dat geval gewijzigde regelgeving en het veranderde bewustzijn over voedselveiligheid ook tot de waargenomen resultaten kunnen hebben geleid. Dergelijke rivaliserende verklaringen kunnen alleen worden uitgesloten als deze verandering in een controlegroep niet worden waargenomen.

## 2.5 Objecten van toezicht

In een deel van de onderzochte studies zijn anonieme, individuele burgers de onder toezicht gestelden (studies cameratoezicht, 1, 2 en 3; verkeersstudies, 13 - 33; drugsoverlast, 5). In een aantal studies zijn de normadressaten bedrijven en instellingen (scholen, 34 - 37; milieutoezicht, 9 - 12; restaurants, 38 - 44).

Het is de vraag of dit onderscheid van belang is voor de wijze waarop gedragsveranderingen van de onder toezicht gestelden moeten worden gemeten. Dat lijkt niet vanzelfsprekend te zijn. De effectiviteitsonderzoekers die er in slagen een causale relatie te leggen met de maatschappelijke effecten, richten zich vaak op statistisch materiaal - veelal uit bestaande databestanden. Dat is zowel het geval bij toezichtsinterventies die individuele burgers als onder toezicht gestelde hebben, als bij interventies die zich richten op bedrijven en instellingen.

Illustratief is het al aangehaalde onderzoek naar de effecten van *naming and shaming* op het gedrag van de onder toezicht gestelde spaarbanken. Het onderzoek betreft de reactie van rekeninghouders op de bekendmaking van negatieve informatie over banken (studie 45). Het doel van *naming and shaming* is dat de banken de regels beter gaan naleven. Dat doel wordt, volgens de beleidstheorie, echter bereikt over de band van klantengedrag. *Naming and shaming* is dan alleen effectief ter bevordering van naleving indien de klanten de beoogde reactie vertonen. De meting van klantreacties wijst dan uit of de onder toezicht gestelde banken de gewenste gedragsreactie vertonen. Een dergelijke indirecte

meting van effectiviteit komt vaker voor. Zo is bij de onderzoeken naar het effect van voedselinspecties in restaurants als effectvariabele een aantal keren het vóórkomen van gevallen van voedselvergiftiging gemeten (studies 40 en 42). En het onderzoek naar de effectiviteit van schoolinspecties vindt plaats door een causaal verband te leggen met de prestaties van leerlingen (onderzoeken 34 – 37).

Deze studies hebben gemeenschappelijk dat nagegaan wordt of er een causaal verband aanwezig is tussen de interventie en de verandering van het gedrag door een indirecte meting van veranderingen van kenmerken van rekeninghouders (bij banken), consumenten (bij restaurants) en leerlingen (bij scholen). Dat onderzoekers die keuze maken is goed te begrijpen. Op die manier kan immers ook wanneer zich slechts beperkte effecten voordoen een causaal verband worden aangetoond; bij een groot aantal waarnemingen kan dan al sprake zijn van significantie.

## 2.6 Samenstelling onderzoeksgroep en controlegroep

Voor een hoge interne validiteit vereist een effectiviteitsmeting een experimentele groep en een controlegroep. Maar hoe worden dergelijke groepen samengesteld? Onderzoek is van hoog niveau wanneer beide groepen op aselechte wijze worden samengesteld. Op die manier worden systematische verschillen immers uitgesloten. In de praktijk is het aselekt samenstellen van de onderzoeksgroep, maar ook van de controlegroep, vaak onmogelijk. Maar soms is dat anders.

In de studies over de effecten van verkeersveiligheidsmaatregelen wordt veelal aannemelijk gemaakt dat verkeersdeelnemers zich at random verdelen over het onderzoeksgebied en het controlegebied. In die studies is er vaak wel een ander probleem en dat is dat ook andere omstandigheden rond de toezichtsinterventie in de vergelijking tussen experimentgroep en controlegroep constant moeten zijn. Als het al lukt de controlegroep en de experimentgroep op identieke wijze samen te stellen, dan moeten ook de omstandigheden waaronder de waarnemingen plaatsvinden gelijk zijn. Dat is vaak één van de lastige punten bij onderzoek naar het effect van verkeersveiligheidsmaatregelen. Wanneer het effect van handhaving van de maximumsnelheid met camera's op een bepaald traject wordt gemeten, moet als controle onderzoek worden gedaan op een vergelijkbaar traject. Wil sprake kunnen zijn van een vergelijkbaar traject, dan moeten fysieke kenmerken vergelijkbaar zijn (hoogteverschillen, bochten etc.), maar ook de verkeersintensiteit moet overeenkomen. Een ander punt is dat controlegebied en experimentgebied als het gaat om verkeersmaatregelen een vergelijkbare mate van normschendingen moeten laten zien voorafgaand aan de interventie. In het onderzoek 'Evaluation of a high visibility enforcement project focused on passenger vehicles interacting with commercial vehicles' (studie 20) werden om die reden wegen geselecteerd met een relatief hoge mate van niet-naleving. Bovendien werd rekening gehouden met het aantal voertuigen dat dagelijks de weg gebruikte. Slagen onderzoekers er in op dergelijke kenmerken te controleren, dan nog kunnen zich vertekeningen voordoen, bijvoorbeeld omdat gedurende de meting de verkeersintensiteit op het ene traject sterker groeit dan op het andere traject.

In een onderzoek van de Campbell Collaboration Group (studie 1) naar cameratoezicht zijn alleen gebieden geselecteerd waar ten minste twintig keer een interventie nodig was geweest in het verleden. Een minimum van twintig incidenten in de voorperiode was gewenst omdat anders de uitkomsten misleidend kunnen zijn en veranderingen in crimineel gedrag op de desbetreffende plek te klein zouden zijn om duidelijke resultaten op te kunnen leveren.

De minimale omvang van het aantal waar te nemen incidenten is een van de omstandigheden die ertoe kan nopen om experimentele en controlegroepen te gebruiken waarbij de selectie van de onderzoekseenheden niet at random heeft plaatsgevonden. In algemene zin geldt dat hoe kleiner het aantal te bestuderen normadressaten, hoe beter de variatie in de totale populatie in kaart moet worden gebracht voor de selectie van de te bestuderen gevallen. Een probleem daarbij is dan weer dat de kenmerken van de situaties waarin het gedrag wordt bestudeerd tevoren vaak onvoldoende bekend zijn, zodat een keuze niet aselekt kan worden gemaakt. Een benadering die dikwijls wordt gevolgd om dit probleem te ontlopen is het bestuderen van een groot aantal gevallen, zodat het aannemelijk is dat de variatie voldoende in beeld wordt gebracht. Een voorbeeld kan worden ontleend aan de effectiviteitsstudie van de VWA over de naleving van hygiënevoorschriften in Chinese (afhaal)restaurants (onderzoek 38). Door de controlegroep samen te stellen uit alle restaurants uit de omgeving Rotterdam en de onderzoeksgroep uit de restaurants uit de omgeving Den Haag, is een voldoende adequate onderzoeksopzet tot stand gekomen.

Doorgaans is het streven er op gericht experimentele groepen en controlegroepen zoveel mogelijk op elkaar te doen lijken. Als gerandomiseerde toewijzing niet lukt is het op één na beste design dat van *matching pairs*: elke eenheid in een experimentele groep heeft een precies vergelijkbare eenheid in de controlegroep. Dit niveau kan onder niet-experimentele omstandigheden zelden bereikt worden. Een fraai voorbeeld waarin dat wel is gelukt kan ontleend worden aan recent criminologisch onderzoek.<sup>6</sup> In dat project werd de kans op recidive na een gevangenisstraf vergeleken met de recidive na een werkstraf. Van de gestraften is eerst een aantal cruciale variabelen beschreven (zoals de criminele geschiedenis, het type delict en persoonskenmerken zoals leeftijd, geslacht en herkomst). Vervolgens zijn beide groepen zodanig samengesteld dat tussen de groepen geen significante verschillen meer bestonden. Daarna is gekeken naar de kans op recidive. Alternatieve verklaringen voor verschillen in recidive zijn zo uitgesloten.

In de bestudeerde onderzoeken wordt doorgaans gebruik gemaakt van vergelijkbare groepen of gebieden, waarbij zoveel mogelijk gecontroleerd wordt voor cruciale variabelen. Bij het onderzoek naar de effecten van het cameratoezicht dat gehouden is op het Stationseiland in de gemeente Amsterdam (studie 2) was dat niet het geval. Het hele district werd gebruikt als controlegebied voor het veel kleinere cameragebied. Het voordeel hiervan was dat daarmee de trend in het cameragebied kon worden afgezet tegen brede-

---

<sup>6</sup> Hilde Wermink, Arjan Blokland, Paul Nieuwbeerta & Nikolaj Tollenaar, Recidive na werkstraffen en na gevangenisstraffen. Een gematchte vergelijking. *Tijdschrift voor Criminologie*, 2009 (51), 3, p. 211-227.

re trends in de ontwikkeling van de criminaliteit. In het onderzoek waren de gebieden wel vergelijkbaar in bij voorbeeld de mate van criminaliteit in het gebied.

Toch zijn er in de gevonden onderzoeken ook voorbeelden waarbij onderzoeksgroep en experimentgroep volgens een toevalsverdeling zijn samengesteld. Die onderzoeken bevinden zich dan op MSMS-5 niveau. In het Amerikaanse onderzoek naar effecten van een handhavingsprogramma gericht op het tegengaan van de verkoop van tabak aan jongeren (studie 61), zijn 319 verkoopadressen ingedeeld in zes paren van plaatsen in Erie County in de staat New York. Het toeval bepaalde of de verkoopadressen aan de interventie blootgesteld zouden worden. Opvallend is dat de verkoopadressen bij geconstateerde naleving van de verbodsbepaling een felicitatiebrief ontvingen! Overigens werd geen significant verschil in naleving tussen de onderzoeksgroep en de controlegroep gevonden. De verklaring daarvoor was dat de bezochte verkoopadressen veelal behoorden tot een keten van winkels. Ook winkels die niet aan de interventie blootgesteld werden, werden via collega's daarover wel geïnformeerd, hetgeen ook daar leidde tot de gewenste gedragsreactie. Een goed voorbeeld van *spill over*, ofwel gewenste, niet beoogde neveneffecten! Hoewel dus sprake was van een onderzoek op MSMS-5 niveau, was het design op een andere manier weer gebrekkig, waardoor geen causaal verband aangetoond kon worden. In een ander experiment uit dezelfde periode (1994/95) in Central Harlem, New York (studie 62) werden eveneens op gerandomiseerde wijze winkels aan een controlegroep of een experimentgroep toegewezen. In dit experiment bleek wel dat de winkels die met toezichtsinterventies werden geconfronteerd een substantiële daling van overtredingen te zien gaven in vergelijking met de winkels waarvoor dat niet gold.

## 2.7 Soorten toezichtsinterventies

De bestudeerde onderzoeken laten een grote variëteit aan toezichtsinterventies zien. Dat gaat van voorlichting en communicatie, tot scholing en training en van boeteoplegging tot massamediale campagnes. Vaak is het voorwerp van onderzoek het bezoek van de inspecteur aan de school, het restaurant, de werkplaats of de transportonderneming. Bij verkeerstoezicht is de interventie de *directe waarneming* van het verkeersgedrag door een camera of een surveillance-eenheid. Soms is sprake van bijzondere manieren van toezichtsuitoefening. In de Amerikaanse studie over de handhaving van het verbod tot verkoop van tabak aan minderjarigen (studie 61) werden minderjarigen geïnstrueerd tabak te kopen. Wanneer de winkelier aan de minderjarige tabak verkocht, kwam vervolgens de inspecteur de zaak binnen. De Amerikaanse (studie 69) en de Australische (studie 70) belastingdienst stuurden *brieven* aan belastingplichtigen *met* in wisselende *normatieve* bewoordingen gestelde *boodschappen*. Op die manier werd de hypothese getoetst dat daardoor de aangiftebereidheid zou veranderen. Die hypothese werd overigens in beide onderzoeken verworpen.

Het Centraal Planbureau deed onderzoek gedaan naar de effecten van *schoolbezoeken* van inspecties aan basisscholen (studie 34). De analyse is gebaseerd op een bestand van alle Cito-scores in Nederland over de jaren 1999-2003. Er wordt in eerste instantie onderscheid gemaakt tussen de scholen die niet zijn bezocht door de inspectie en scholen die

wel zijn bezocht. In de analyse wordt vervolgens ook onderscheid gemaakt tussen kortere en intensievere bezoeken aan scholen.

In een onderzoek uit de Filippijnen naar de populatiedichtheid van schelpdieren in een beschermd gebied (studie 9) is gekeken naar het effect van het afsluiten van een bepaald perceel zeewater voor economische activiteiten en het plaatsen van uitkijktorens en het monitoren van visserijactiviteiten vanuit die torens op de toename van het aantal en de grootte van schelpdieren in een gebied. Twee gebieden werden met uitkijktorens uitgerust; in twee andere gebieden stonden geen uitkijktorens. Het is een zeer passieve vorm van toezicht waarbij de aanwezigheid van de uitkijktorens er voor moesten zorgen dat de beschermde gebieden met rust werden gelaten.

In de Verenigde Staten is een evaluatie gedaan van een verkeersveiligheidsproject waarbij het doel was dat het toezicht op zoveel mogelijk manieren zichtbaar moest zijn. Hiervoor werd het toezicht aangekondigd door media, zoals televisie, radio en krantenadvertenties en posters, banners, flyers en aankondigingen op de weg. Ook reden er tijdens de interventieperiodes vrachtwagens op de wegen waarop toezicht werd gehouden met aankondigingen van het toezicht (studie 20).

Een intensief communicatietraject was ook onderdeel van het Canadese onderzoek naar de effecten van toezicht op het dragen van veiligheidsgordels door bestuurders (onderzoek 21). Via onder meer televisie, radio en publicatieborden werd de verplichting tot het dragen van gordels kenbaar gemaakt. Ook werd intensivering van de controles aangekondigd. Na afloop van de handhavingsacties werd op dezelfde manier gecommuniceerd over de effecten van de handhavingsinspanning op de normconformiteit.

In het onderzoek naar cameratoezicht van de Campbell Collaboration (studie 1) zijn 12 gebieden met *cameratoezicht* onderzocht. De camera's werden geplaatst in centra van steden en dorpen en op plekken in gebouwen. De metingen vonden plaats 6, 12 en 24 maanden voor en na de plaatsing van de camera's. De onderzoekers hebben gebruik gemaakt van de geregistreerde misdadcijfers die relevant zijn voor het onderzoek.

In een ander onderzoek naar drugshandhaving (studie 4) is de effectiviteit van verschillende manieren van toezicht door de politie onderzocht, waarbij er altijd een vergelijking in tijd en een controlegroep aanwezig was. In het onderzoek zijn verschillende manieren van toezicht door de politie toegepast en is een vergelijking ook in verschillende methodes mogelijk. De verschillende toezichtsinterventies die zijn onderzocht waren probleemgeoriënteerd patrouilleren, patrouilleren bij zogenaamde hotspots (gebieden met veel drugsoverlast) en patrouilles in de hele buurt. Door deze verschillende interventies en de effecten daarvan te vergelijken was het mogelijk een afweging te maken tussen de kosten en de baten van de toezichtsinterventies. Het onderzoek bracht aan het licht dat patrouilleren in een klein gebied uiteindelijk net zo effectief is als in de hele wijk.

Meer in het algemeen valt op dat het toezichtsoptreden in veel van de onderzochte studies een sterk communicatief aspect heeft en dat regelmatig een combinatie van interventies wordt toegepast waarvan communicatie deel uitmaakt. Een reden zou kunnen zijn

dat juist effecten van communicatie relatief goed te onderzoeken zijn, niet in de laatste plaats omdat daarmee veel ervaring is opgedaan op andere terreinen zoals marketing.

Het toepassen van een combinatie van toezichtsinterventies heeft overigens ook nadelen. In het onderzoek naar de naleving van hygiënevoorschriften door de Chinese horeca (studie 38) werden meerdere interventies toegepast op de onderzoeksgroep. In de onderzoeksgroep werd gehandhaafd als niet volgens de hygiënevoorschriften werd gewerkt. Daarnaast vond ook in de restaurants behorend tot de controlegroep een intensief communicatietraject plaats. Personeel van de betreffende restaurants werd uitgenodigd voor voorlichtingsbijeenkomsten, brochures over voedselveiligheid en over de toepassing van de hygiënevoorschriften werden uitgereikt en het keukenpersoneel van de restaurants in de onderzoeksgroep werd in de gelegenheid gesteld deel te nemen aan trainingen over voedselveiligheid. De VWA concludeert op basis van de resultaten van de effectiviteitsmeting dat de inzet van de handhaving succesvol is. Strikt genomen kan dat echter niet uit het onderzoek worden geconcludeerd: in de controlegroep was immers geen sprake van handhaving en evenmin van het uitgebreide communicatietraject. Welk effect elk van de toezichtsinterventies afzonderlijk had kan dus per interventie niet worden vastgesteld.

Ook in het onderzoek naar de effectiviteit van cameratoezicht op het stationseiland (project 2) is het niet gelukt de toezichtsinterventie door middel van cameratoezicht te 'isoleren' van andere interventies die gelijktijdig plaats vonden. En hoewel er een controlegebied was, is dit zodanig samengesteld dat er geen zuiver 'met/zonder' design tot stand kwam.

## 2.8 Verschillende methoden van dataverzameling

In de bestudeerde onderzoeken is een veelheid aan methoden van dataverzameling aan de orde. Wat opvalt is dat in veel onderzoeken gebruik wordt gemaakt van bestaande databestanden. Dat databestanden informatie geven over de onafhankelijke variabele, het politieoptreden of het inspectiebezoek is niet verrassend. Maar opvallend is dat zo vaak gebruik wordt gemaakt van databestanden die informatie bevatten over de door de interventie te beïnvloeden afhankelijke variabele. Politieregistraties vertellen ons het nodige over de maatschappelijke effecten en/of de nalevingseffecten, zie bij voorbeeld de informatie die in de onderzoeken wordt gebruikt over criminaliteit (studies 1 t/m 4), verkeersovertredingen en ongevallencijfers (studies 13 t/m 33). Ook bij inspecties zelf zijn er veel databestanden, zie bijvoorbeeld de studies over onderwijsprestaties (studies 34 t/m 36), arbeidsongevallen (studies 63 t/m 67) en de belastingdienst (studies 68 t/m 70). De gezondheidszorg levert informatie over verkeersslachtoffers en –doden en over voedselvergiftigingen (studies 38 t/m 44).

In veel van de verzamelde studies worden 'nieuwe' data verzameld. Heel vaak gaat het dan om data die aanvullend zijn ten opzichte van de bestaande databases die informatie verschaffen over de causale relatie tussen de interventie en de effecten daarvan. Die data zijn bij voorbeeld bedoeld om die relatie te kunnen interpreteren. In verschillende van de verkeersonderzoeken is met die bedoeling *aanvullend enquêteonderzoek* gehouden onder verkeersdeelnemers. Soms was dat onderzoek bedoeld om te kunnen controleren of de

onderzoeksgroep en de controlegroep op vergelijkbare wijze waren samengesteld, soms was het bedoeld om te achterhalen hoe een causale relatie precies verklaard moet worden (studies 15, 20 en 23). Uit Amerikaans onderzoek uit de jaren negentig naar het effect van toezicht op het dragen van veiligheidsgordels (studie 23) blijkt bij voorbeeld dat de introductie van de verplichting tot het dragen van veiligheidsgordels weliswaar heeft geleid tot een hoger nalevingspercentage en een vermindering van ongelukken, maar dat die verandering niet bij alle weggebruikers even hoog was. De bestuurders met de grootste kans op een ongeluk – jonge, mannelijke chauffeurs – reageerde het slechtst op de maatregel. Een dergelijke combinatie van onderzoeksmethoden levert dus interessante informatie op die de toezichthouder er toe kan bewegen zijn communicatie- en handhavingsstrategie op een specifieke doelgroep af te stemmen.



# Handreikingen

## 3.1 Inleiding

Bovenstaande analyse laat zien dat het moeilijk, maar niet onmogelijk is om toezichtsinterventies meetbaar te maken. Vooral de meting van maatschappelijke effecten van toezicht is niet eenvoudig. Effectiviteitsonderzoek stelt hoge eisen aan het onderzoeksdesign, zo is uit het bovenstaande inmiddels duidelijk geworden. Maar vooral werd duidelijk dat het doen van effectiviteitsonderzoek gewoon mogelijk is. Op basis van de voorgaande analyse worden in dit hoofdstuk handreikingen geformuleerd die de Nederlandse toezichthouders kunnen gebruiken bij het opzetten van onderzoek naar de effectiviteit van hun toezichtsinterventies.

Het eerste advies dat in paragraaf 3.2 wordt verwoord gaat over het eenvoudig houden van het onderzoeksdesign. Vervolgens gaan we in paragraaf 3.3 in op het belang van een goede analyse van de toezichtsstrategie voor het uitvoeren van effectiviteitsonderzoek. Onderzoek naar de effectiviteit van toezicht is ‘serieus’ onderzoek. Daarom het advies dergelijk onderzoek alleen uit te voeren als onderdeel van een bestaande toezichtsstrategie (3.4). Verdere aanbevelingen hebben betrekking op het doen van kwalitatief onderzoek (3.5), het gebruiken van metingen in een tijdreeks (3.6), de samenstelling van de controlegroep (3.7) en het belang van neveneffecten (3.8).

## 3.2 Hou het eenvoudig

De vele voorbeelden die in het vorige hoofdstuk de revue passeerden, maakten duidelijk dat onderzoek om de effectiviteit van toezichtsinterventies vast te stellen, niet eenvoudig is. Effectiviteitsonderzoek is kostbaar en tijdrovend. Daarbij is ook het onderscheid tussen nalevingseffecten en maatschappelijke effecten aan de orde gesteld. Wanneer toezichthouders (nog) geen enkele ervaring hebben met het (laten) uitvoeren van effectiviteitsonderzoek is het aan te bevelen te starten met een zo eenvoudig mogelijk onderzoeksdesign.

Er zijn verschillende manieren om het design eenvoudig te houden. In de eerste plaats kan ervoor worden gekozen het onderzoek te richten op het effect dat de toezichtinspanning had op de naleving van de gestelde norm. In het vorige hoofdstuk zijn vele voorbeelden gegeven van onderzoek dat de nadruk legt op naleving. Het onderzoeken van maatschappelijke effecten van toezicht is in de meeste gevallen een ingewikkelder klus, zo werd duidelijk. Beginnen met onderzoek van nalevingseffecten heeft bovendien vaak als voordeel dat de extra stap, het in kaart brengen van de maatschappelijke effecten, desgewenst dikwijls daarna nog gezet kan worden. Het onderzoek dat in het overzicht in bijlage 2 is gepresenteerd op het terrein van verkeersveiligheid biedt daarvan een goede illustratie. Veel van dat onderzoek gaat over maatschappelijke effecten en heeft ongevals cijfers, het aantal doden en (zwaar)gewonden, als afhankelijke variabele van de toezichtsinterventie gemeten. Maar er zijn ook verkeersveiligheidsstudies die zich richten op de naleving van bepalingen over maximumsnelheid, het dragen van autogordels of het verbod op het rijden na alcoholgebruik. Bij die laatste studies is het meestal mogelijk – zij het soms na bewerkingen van het cijfermateriaal – als vervolgstap de relatie tussen de verandering van de naleving en de veranderingen in de ongevalscijfers in kaart te brengen.

*Toezichthouders die ervaringen willen opdoen met effectiviteitsonderzoek doen er verstandig aan te beginnen met het doen van onderzoek naar de effecten van het toezicht op de naleving.*

Een tweede mogelijkheid om de kosten en de benodigde tijd voor het doen van effectiviteitsonderzoek te beperken is gebruik te maken van bestaande databestanden. In onderzoek op het terrein van arbeidsomstandigheden bij voorbeeld werden cijfers over (fatale) arbeidsongevallen gehanteerd als indicator van het toezichtseffect. Veelal zijn dergelijke cijfers bij werkgevers bekend, maar in enkele onderzoeken bleek dat materiaal ook in de registraties van de toezichthouder aanwezig te zijn. Andere vergelijkbare variabelen die relatief eenvoudig beschikbaar zijn omdat ze in bestaande databases worden geregistreerd, zijn bijvoorbeeld Cito-scores, ongevalscijfers, cijfers over de handel in aandelen, belastingafdrachten. In hoofdstuk 2 is veel onderzoek besproken waarin dergelijke cijfers in een effectiviteitsonderzoek zijn gebruikt door daarop een secundaire analyse te plegen. Activiteiten van de toezichthouder zelf – zoals het aantal bezoeken aan scholen, het aantal gemeten snelheidsovertredingen, het aantal geïnspecteerde voertuigen, bezochte restaurants – worden in ieder geval goed geregistreerd. In een relatief eenvoudig onderzoeksdesign kunnen dergelijke, reeds bestaande databestanden worden gebruikt voor een eerste verkenning van het bestaan van een (causaal) verband. Desgewenst kan vervolgens alsnog de keuze worden gemaakt voor het inzetten van aanvullende methoden van dataverzameling, met het oog op het interpreteren van die verbonden.

*Inventariseer de beschikbaarheid en bruikbaarheid voor effectiviteitsonderzoek van alle bestaande gegevensbestanden voordat zelf data worden verzameld.*

Een derde manier om het onderzoeksdesign eenvoudig te houden is te beginnen met het onderzoeken van eenvoudige verbanden. In plaats van de effectiviteit van de interventies van de toezichthouder op het terrein van verkeer en vervoer in het algemeen, kan gekozen worden voor de effectiviteit van bepaalde interventies (snelheidsmetingen, media-campagnes, voertuiginspecties) op bepaalde verkeersdeelnemers (een automobilist, een vrachtvervoerder, een taxibedrijf). De keuze voor een eenvoudiger design maakt het eerder mogelijk uitspraken te doen over de effecten van de interventie. Het overzicht geeft mooie voorbeelden van dergelijk onderzoek, zoals het onderzoek van de Nederlandse Inspectie Verkeer en Waterstaat naar de veiligheid van rolstoelvervoer. Door op die manier stap voor stap ervaring op te bouwen met effectiviteitsonderzoek is de toezichthouder in staat desgewenst een steeds geavanceerder en complexer design te kiezen.

Een lastige kwestie is dat een toezichtsinterventie zelden alleen komt. Een toezichthouder die graag wil weten wat de effecten zijn van de inzet van een bepaalde interventie, krijgt daardoor vaak slechts informatie over de causale relatie met de mix van interventies. Een voorbeeld daarvan, besproken in paragraaf 2.7, is het Nederlandse VWA-onderzoek (studie 38), dat de effecten van een combinatie van interventies (cursussen, communicatie, handhaving) gericht op Chinese restaurants, in kaart bracht. In het onderzoeksrapport werden uitspraken gedaan over de effectiviteit van één van de interventies, die strikt genomen niet op de bevindingen van het onderzoek konden worden gebaseerd. Bij het opzetten van het onderzoek dient men hiermee al rekening te houden.

*Ervaring met effectiviteitsonderzoek moet stap voor stap worden opgebouwd, beginnend met een eenvoudig design, waarbij wordt getracht een enkelvoudige toezichtsinterventie te relateren aan het gedrag van een specifieke groep normadressaten.*

### 3.3 Formuleer verwachtingen

Effectiviteitsonderzoek is erop gericht om relaties vast te stellen tussen een of meer onafhankelijke variabele(n) (toezichtseffecten) en een of meer afhankelijke variabele(n) (waaronder toezichtsactiviteiten) waarbij een voor/na meting wordt uitgevoerd en een vergelijking wordt gemaakt met een controlegroep (met/zonder benadering). De aanpak van een toezichthouder is doorgaans gebaseerd op veronderstellingen over de mogelijke relaties tussen de variabelen die worden onderzocht: de beleidstheorie van de toezichthouder. Goed effectiviteitsonderzoek begint daarom met een analyse van de bestaande

toezichtsstrategie en met de reconstructie van de beleidstheorie die aan de toezichtsstrategie ten grondslag ligt.

De reconstructie van de beleidstheorie leidt tot het ontwikkelen van hypothesen die in het onderzoek worden getoetst. In die hypothesen wordt een verband gelegd met de doelstelling die de toezichthouder met zijn interventie wil realiseren en de wijze waarop het effect tot stand komt dat aan die doelrealisatie bijdraagt. De beleidstheorie werkt de causale paden uit volgens welke de effecten tot stand komen die de interventie idealiter moet opleveren. In het onderzoek wordt als het ware getest of die veronderstelde causale paden op juiste aannames berusten. Beschikt de toezichthouder niet zelf over een uitgeschreven beleidstheorie, dan kan de onderzoeker die theorie reconstrueren. Een voorbeeld is de reconstructie van de beleidstheorie die Ehren in haar eerder genoemde proefschrift maakte van de Wet Onderwijstoezicht (WOT).

Een alternatief voor bovenstaande werkwijze is dat de evaluatieonderzoeker genoeg neemt met het empirisch onderzoeken van één of meer verwachtingen van de toezichthouder over de wijze waarop het gedrag van de normadressaat wordt beïnvloed door de interventie.

Bij deze exercitie moet de toezichthouder ook nadrukkelijk proberen zich tevoren een beeld te vormen van gewenste of ongewenste neveneffecten.

*Bij het ontwikkelen van een onderzoeksdesign dienen onderzoekers de beleidstheorie van de toezichthouder te hanteren. Bij het ontbreken daarvan, wordt die beleidstheorie gereconstrueerd. Een vereenvoudigd alternatief daarvoor is het expliciteren van één of meer verwachtingen van de toezichthouder over de effecten van zijn interventie.*

### 3.4 Inbedding in de toezichtsstrategie

In de verzamelde studies komt een enkele keer een onderzoek voor naar effecten van een toezichtsinterventie die uitsluitend plaats vond in het kader van dat onderzoek. Kennelijk was de gedachte er een van 'trial and error': we proberen eens wat uit. In andere onderzoeken is weliswaar sprake van permanent toezicht op de normadressaat, maar is de onderzochte mix van interventies wel exclusief voor de bedoelde onderzoeksperiode gehanteerd.

Effectiviteitsonderzoek is te duur en te tijdrovend voor dergelijke proefopstellingen. Ook experimenten moeten worden ingebed in een 'normaal' toezichtsarrangement. In de eerste plaats omdat de onderzoeksresultaten anders weinig relevantie hebben. In de tweede plaats omdat niet-ingebed onderzoek ten koste gaat van de geloofwaardigheid van effectiviteitsonderzoek. Het is dan immers de vraag of de in het onderzoek in kaart gebrachte positieve effecten op de lange(re) termijn wel beklijven. Is een gedragsverandering of een maatschappelijke effect van blijvende aard wanneer de onderzochte interventie wordt

gestaakt? Hier kunnen effectmetingen strikt genomen geen uitspraken over doen omdat de resultaten slechts gelding hebben binnen het onderzochte design waar die interventies onderdeel van uitmaakten.

*Effectiviteitsonderzoek moet zoveel als mogelijk is realistische toezichtssituaties evalueren. Ook experimenten dienen te zijn ingebed in een bestaande toezichtsstrategie.*

In het verlengde daarvan geldt dat onderzoekers zich bewust moeten zijn van mogelijke *onderzoekseffecten*. Niet altijd kan met volledige zekerheid worden gezegd dat het uitvoeren van het onderzoek naar de effecten niet zelf heeft bijgedragen aan het ontstaan van positieve (of negatieve) maatschappelijke of nalevingseffecten. Metingen worden immers niet gedaan in een laboratorium waar externe invloeden kunnen worden uitgeschakeld. Dit heet ook wel het ‘Hawthorne’ effect: het feit dat het goed uitgewerkte design van het effectiviteitsonderzoek zelf ook bijdraagt aan de normnaleving kan niet worden verwaarloosd. Dergelijke effecten doen zich vooral voor bij experimenten, waarbij in een ‘proefopstelling’ geprobeerd wordt een bepaalde interventie te koppelen aan een bepaald gewenst effect. De normadressaat kan er zich zodanig van bewust zijn dat wordt ‘meegekeken’ dat hij op basis daarvan zijn gedrag wijzigt. Onderzoekers moeten er voor waken dat onderzoek wordt gedaan in ‘modelmatige’ omstandigheden die niet realistisch zijn. Experimenten moeten realistisch zijn – toezichtsinterventies die worden uitgetoetst moeten ook daadwerkelijk toepasbaar zijn.

*Onderzoeksactiviteiten kunnen de causale relatie tussen een toezichtsinspanning en de effecten daarvan beïnvloeden. Dat pleit voor het doen van onderzoek naar daadwerkelijk toepasbare – en herhaalbare – toezichtsinterventies.*

### 3.5 Doe ook kwalitatief onderzoek

Het is dikwijls niet eenvoudig van de effecten van toezicht vast te stellen of die significant zijn. Statistisch significant wil zeggen dat het verschijnsel niet bij toeval is opgetreden, en zich bij herhaling van het experiment of de waarneming zeer waarschijnlijk opnieuw zal voordoen. Onderzoekers van de effectiviteit van toezichtsinterventies kiezen er vaak voor grote aantallen gedragingen te bestuderen omdat de aldus verzamelde gegevens zich beter lenen voor statistische analyse dan data van een klein aantal onderzoekseenheden. Dat verklaart dat veel effectiviteitsonderzoek wordt gevonden op beleidsterreinen waar sprake is van grote aantallen, zoals op het terrein van verkeersveiligheid en voedselveiligheid, waar tevens databestanden aanwezig zijn die een meting van dergelijke gedragsveranderingen goed mogelijk maken.

Het in dit onderzoek gehanteerde selectiecriteria MSMS-3+ leidde er bijna vanzelf toe dat overwegend kwantitatieve studies naar de effectiviteit van toezichtsinterventies werden gevonden. Dat neemt echter niet weg dat ook onderzoek dat langs meer kwalitatieve weg de aannemelijkheid van bepaalde causale samenhangen aantoonde waardevol kan zijn. Toezichthouders moeten ook oog hebben voor de mogelijkheden van kwalitatieve gegevensverzameling over de effectiviteit van toezicht. Een *verklarend casestudy-design* kan bijvoorbeeld zeer waardevolle informatie opleveren.<sup>7</sup> In verschillende van de verzamelde studies is sprake van het (aanvullende) gebruik van kwalitatieve methoden van gegevensverzameling, zoals interviews en observaties. Op die manier kan extra duiding aan gevonden verbanden worden gegeven.

*Voor het vaststellen van verbanden tussen toezichtsinterventies en de effecten daarvan is men niet uitsluitend aangewezen op kwantitatief onderzoek en statistische analyse. Bij verklarend effectiviteitsonderzoek kan aanvullend goed gebruik worden gemaakt van kwalitatieve methoden van dataverzameling.*

### 3.6 Tijdreeksmeting kan soms ook

Een probleem van effectmetingen is vaak dat informatie nodig is uit een periode voorafgaand aan het toezicht. Die informatie is echter lang niet altijd beschikbaar, of van voldoende kwaliteit. Wanneer een toezichthouder besluit een effectiviteitsonderzoek uit te voeren wil hij graag snel resultaat en ontbreken vaak de tijd en de mogelijkheden om de situatie *zonder* interventie in kaart te brengen. Een alternatief dat veelal wordt gevolgd is het doen van metingen in een tijdreeks. Daarbij wordt aangenomen dat gedrag niet ineens verandert, maar zich in de tijd ontwikkelt. Onder de gevonden studies zijn er verschillende waarin een *ex-ante* meting ontbreekt, maar waarbij als alternatief gekozen is voor metingen op meerdere tijdstippen (studies 17, 35, 42, 59, 60, 65).

*Voor het aantonen van een causale relatie tussen een toezichtsinterventie en de effecten daarvan wordt idealiter een voor/na meting gecombineerd met een met/zonder benadering. Wanneer het niet mogelijk is een voormeting uit te voeren kan worden gekozen voor het uitvoeren van metingen in een tijdreeks.*

Dergelijk longitudinaal onderzoek moet wel worden aangevuld met metingen bij een controlegroep, om uit te sluiten dat de opgespoorde effecten moeten worden toegeschreven aan andere factoren dan de toezichtsinterventie(s).

<sup>7</sup> N. Struiksmā, J. de Ridder & H.B. Winter, *De effectiviteit van bestuurlijke en strafrechtelijke handhaving van milieuwetgeving*. Reeks onderzoek en beleid, nr. 253, Den Haag, BJU: 2007.

Over de kwestie van het vergelijkbaar samenstellen van de controlegroep en de onderzoeksgroep is hiervoor in paragraaf 2.6 reeds het nodige opgemerkt. Idealiter is sprake van een toewijzing aan de onderzoeksgroep (experimentgroep, programmagroep) en de controlegroep op basis van het toeval. In dat geval kan het onderzoek op MSMS-5 niveau worden geschaald. 'Next best' is dat de onderzoekers bij de keuze van onderzoekseenheden zodanig selecteren dat men kan controleren voor bepaalde variabelen. Bij onderzoek naar verkeersveiligheid bij voorbeeld worden het onderzoeks- en het controletracé gekozen op basis van overeenkomsten op zoveel mogelijk cruciale variabelen, zoals fysieke kenmerken, verkeersintensiteit, landelijke of stedelijke omgeving etc. Op die manier worden interveniërende variabelen, die de causale samenhang kunnen verklaren, zoveel mogelijk uitgeschakeld. Gevonden verschillen zijn dan in elk geval niet veroorzaakt door verschillen in deze omstandigheden. De kans dat het gevonden verschil een effect is van de bestudeerde toezichtsinterventie is dan aanzienlijk.

*Bij het uitvoeren van een onderzoek waarbij een met/zonder benadering wordt gevolgd worden normadressaten idealiter at random aan de onderzoeksgroep en de controlegroep toegewezen. Als dat niet mogelijk is, dient bij de keuze en samenstelling van de controlegroep te worden gecontroleerd voor alternatieve verklaringen door overeenkomsten met de onderzoeksgroep te vinden op zoveel mogelijk cruciale variabelen.*

Bij het samenstellen van een controlegroep bestaan soms ethische en juridische dilemma's waarvoor aandacht wordt gevraagd. Kan een toezichthouder er bij voorbeeld voor kiezen selectief te handhaven, wanneer hij vermoedt dat ernstige maatschappelijke gevolgen kunnen optreden wanneer normadressaten zich niet conformeren aan de wettelijke norm? Een politieke principaal zal zich bij overtuigend bewijs van negatieve maatschappelijke consequenties van niet-handhaven niet snel neerleggen bij de keuze van een handhavende dienst om ten behoeve van een gecontroleerd onderzoeksdesign ten aanzien van een bepaalde groep normadressaten af te zien van handhaving. In de literatuur wordt ook nog wel eens melding gemaakt van een juridisch dilemma: het gevolg van niet-handhaven zou kunnen zijn dat de toezichthouder wordt geconfronteerd met schadeclaims van personen of instellingen die als gevolg van normafwijkend gedrag van anderen schade hebben geleden. De Nederlandse rechtspraak laat zien dat handhavende instanties niet snel in rechte verantwoordelijk kunnen worden gehouden. Om een toezichthouder aansprakelijk te kunnen stellen zou er al sprake moeten zijn van opzettelijke en verwijtbare nalatigheid. Een goed doordachte en gemotiveerde onderzoeksopzet zou nooit een dergelijke nalatigheid mogen opleveren.

*Toezichthouders die effectiviteitsonderzoek wensen uit te voeren moeten zich tevoren vergewissen van ethische, politieke en juridische dilemma's bij het ontwikkelen van een onderzoeksdesign.*

### 3.8 Let op neveneffecten

In hoofdstuk 1 is beknopt gewezen op een kwestie die in de literatuur wel bekend staat als 'moral hazard'. Moral hazard behoort tot de meer algemene categorie van *ongewenste neveneffecten van toezicht*. Toezichtsinterventies kunnen leiden tot ongewenste neveneffecten die zichtbaar gemaakt moeten worden om een goed beeld te kunnen geven van de totale gevolgen van de interventie. Een duidelijk voorbeeld daarvan komt naar voren in het Amerikaanse onderzoek naar de gevolgen van het toezicht op de verplichting autogordels te dragen (studie 23). Die toezichtsinterventie kan leiden tot het gewenste nalevingseffect, waardoor meer automobilisten gordels dragen. Verschillende studies wijzen er op dat dit effect zich daadwerkelijk voordoet. Tegelijkertijd kan sprake zijn van compenserend gedrag doordat diezelfde automobilisten zich door het dragen van autogordels beter beschermd voelen tegen (de gevolgen van) ongevallen, waardoor ze riskanter rijgedrag gaan vertonen. Dat ongewenste neveneffect is in de cijfers over ongevallen waarschijnlijk niet terug te vinden, maar het bestaat wel. Het genoemde onderzoek laat zien dat het ongewenste neveneffect dat – ondanks, of zelfs als gevolg van een betere naleving – optreedt, kan worden opgespoord door de inzet van een aanvullende methode van dataverzameling (in dit geval enquêteonderzoek onder bestuurders). Onderzoekers moeten zich dus altijd bewust zijn van dit verschijnsel en zich bij het ontwikkelen van een design afvragen welke neveneffecten kunnen optreden en hoe die kunnen worden opgespoord.

*Bij onderzoek naar de effecten van toezichtsinterventies moeten onderzoekers zich bewust zijn van mogelijk optredende ongewenste neveneffecten. Door daarmee bij het ontwikkelen van een design rekening te houden kunnen neveneffecten worden opgespoord.*

## Samenstelling begeleidingscommissie

Ron Visser, Universiteit Leiden (voorzitter)  
Inge de Wolf, Onderwijsinspectie  
Sjoerd Goslinga, Belastingdienst  
Ralph Vossen, CCV  
Frank Willemsen, WODC

