

Een voorstel voor de onderzoeksprogrammering Effecten van Toezicht

F.Willemsen, WODC

1 Inleiding

1.1 Aanleiding

Meerjarenprogramma Effecten van toezicht

Bij verschillende overheidsorganisaties leeft de behoefte aan meer wetenschappelijke kennis over de werking van toezicht, handhaving en de naleving van regelgeving. Binnen de Inspectieraad¹ bestaat al enkele jaren de ambitie om meer inzicht te verkrijgen in de effecten van toezicht. Effectmeting staat bij alle rijksinspecties evenwel nog in de kinderschoenen. Om een impuls te geven aan het effectonderzoek heeft de Inspectieraad in 2008 *de werkgroep effectmeting* gestart. De opdracht van de werkgroep is het opstellen van het meerjarenprogramma *'Effecten van toezicht'*. De doelstelling van het meerjarenprogramma is als volgt beschreven:

Het verbeteren van de toezichtspraktijk door het meten van effecten van toezicht, waarbij gebruik wordt gemaakt van resultaten van (wetenschappelijk) onderzoek van die toezichtspraktijk. Het programma heeft een Research en Development (R&D) karakter.

Het programma moet aanzetten tot het daadwerkelijk meten van effecten. Gezien de voorgeschiedenis en de complexiteit van het onderwerp is gekozen voor een vierjarig programma. Het uiteindelijke doel is dat er over vier jaar sprake is van een geborgd systeem van effectmeting bij alle rijksinspecties. Een dergelijk systeem heeft de volgende kenmerken:

- Effectmeting is een integraal onderdeel van de toezichtscyclus van de verschillende rijksinspecties;
- Binnen de verschillende rijksinspecties is binnen de hele organisatie draagvlak voor effectmeting: van management tot inspecteur en medewerker;
- Effectmeting vindt binnen de verschillende inspecties plaats via een (wetenschappelijk) verantwoorde onderzoeksmethodologie.

Om een geborgd systeem van effectmeting bij alle rijksinspecties te bereiken, is het van belang dat het meerjarenprogramma *'Effecten van toezicht'* de volgende vier onderling samenhangende processen in gang zet en houdt:

1. Het verder op gang brengen van de discussie over effectmeting binnen de Inspectieraad;
2. Het ondersteunen van rijksinspecties bij de implementatie van effectmeting in de toezichtspraktijk;
3. Het (stimuleren van het) uitvoeren van (wetenschappelijk) onderzoek op het gebied van effectmeting door en in samenwerking met rijksinspecties, en het gebruik van de onderzoeksresultaten door de inspecties;
4. Het in kaart brengen van de meerwaarde van effectonderzoek voor de samenwerkingsactiviteiten tussen de rijksinspecties (bijvoorbeeld in de domeinen).

¹ In de Inspectieraad werken de 15 rijksinspectiediensten samen aan de vernieuwing van het toezicht vanuit het Rijk. De Inspectieraad is het verband voor samenwerking tussen de rijksinspectiediensten.

Er worden momenteel verschillende activiteiten ondernomen die bijdragen aan de bovengenoemde processen. Eén ervan is het samenstellen van een Onderzoeksprogramma Effecten van Toezicht (EvT) dat we in deze notitie zullen presenteren. Dit onderzoeksprogramma is ondermeer gebaseerd op een literatuuronderzoek dat in opdracht van het WODC (en op verzoek van de werkgroep) is uitgevoerd door de Vakgroep Bestuursrecht en Bestuurskunde van de Faculteit Rechtsgeleerdheid van de Rijksuniversiteit Groningen en bestuurskundig en juridisch onderzoeks- en adviesbureau Pro Facto. Omdat het programma 'Effecten van toezicht' zich als doel heeft gesteld de toezichtspraktijk te verbeteren door gebruik te maken van onderzoek naar de effectiviteit van toezicht, is in deze studie gezocht naar beschikbare (wetenschappelijke) kennis over de relatie tussen handhaving en toezicht, gedrag en de effecten. De literatuurstudie geeft een overzicht van designs en best practices. Er worden 70 (vooral buitenlandse) studies beschreven die zijn ingedeeld in een negental beleidsterrein². Het 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. In het rapport wordt op geaggregeerd niveau gerapporteerd over de gevonden studies in termen van effectiviteit, gebruikte onderzoeksdesign en methoden en technieken. Op basis daarvan formuleren de onderzoekers 'handreikingen' voor de toezichtspraktijk. Het Groningse rapport draagt in belangrijke mate bij aan het in gang zetten van de eerder besproken processen (punt 1 t/m 4). In de onderzoeksprogrammering die we voorstellen bouwen we hier op voort.

Het onderzoeksprogramma Handhaven en Gedrag

Drie jaar voor aanvang van het meerjarenprogramma Effecten van Toezicht is het project Handhaving en Gedrag (HH&G) van start gegaan. We noemen dit project omdat het interessante raakvlakken heeft met EvT. Het project HH&G is gericht op het vergaren van empirische wetenschappelijke kennis met een gedragswetenschappelijk karakter. Het *doel* van het project is tweeledig:

- het toegankelijk maken van bestaande wetenschappelijke kennis over regelnaleving;
- en het veiligstellen van de ontwikkeling van nieuwe kennis.

Centraal staat de vraag *of* en in *welke mate* het nalevingsgedrag van burgers, bedrijven en instellingen wordt beïnvloed door individuele motieven en kenmerken, door de wijze van toezicht en handhaving en door andere (sociale) omgevingsfactoren. Wet- en regelgeving en handhaving(sactiviteiten) maken immers deel uit van een complex van factoren die van invloed zijn op het gedrag van burgers en bedrijven. Om de bijdrage hiervan aan het nalevingsniveau vast te stellen en de werking goed te kunnen begrijpen, dient de positie van deze factoren binnen het geheel aan invloeden van het naleefgedrag van individuen en bedrijven in kaart te worden gebracht.

EvT en HH&G vullen elkaar aan, beiden gaan over handhaving en toezicht en de invloed ervan op burgers, bedrijven en instellingen. Het verschil tussen beide zit hem in het niveau waarop we naar de problematiek kijken en in deze notitie zullen we uitleggen hoe beide programmeringen zich tot elkaar verhouden. Voor beide initiatieven (EvT en HH&G) wordt door het WODC een voorstel gedaan voor een onderzoeksprogramma. In dit document presenteren we het onderzoeksprogramma EvT, het onderzoeksprogramma voor HH&G staat in een separaat document.

1.2 Het belang van een onderzoeksprogramma

Een onderzoeksprogramma is meer dan een lijst met voorgenomen onderzoeken. *Allereerst* dient een programma in onze ogen een raamwerk te bevatten waarbinnen onderzoeken vanuit verschillende disciplines en in een zekere samenhang kunnen worden uitgevoerd. Een onderzoeksprogramma zou idealiter ruimte moeten bieden aan het

² Arbeidsomstandigheden: 5 Belastingen: 3 Economie & Financiën: 10 Onderwijs: 4 Openbare Orde en Veiligheid: 8 Milieu: 4 Verkeer & Vervoer: 21 Volksgezondheid: 7 Voedselveiligheid: 7

uitvoeren van zowel monodisciplinair als ook multidisciplinair onderzoek. Daarnaast dient de in het programma beschreven problematiek herkenbaar te zijn voor zowel wetenschappers als voor de maatschappelijke omgeving (in dit geval de omgeving van inspecties en handhavende organisaties).

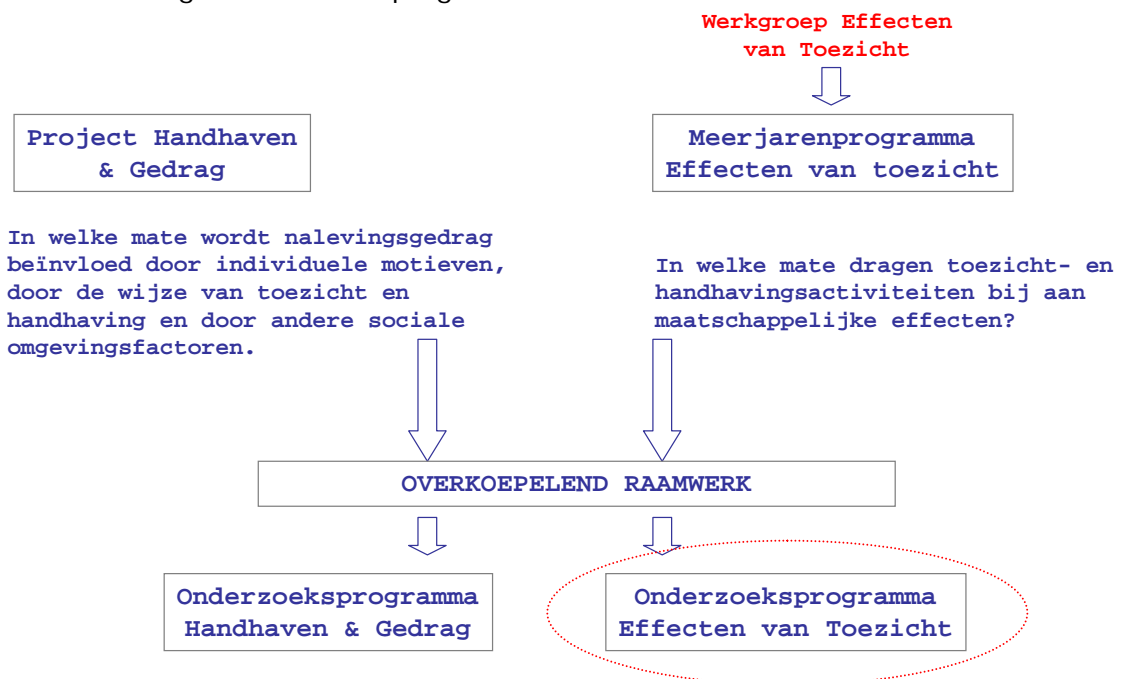
Dat er behoefte is aan het verzamelen en beschikbaar stellen van wetenschappelijke kennis op het gebied van effecten van toezicht, handhaving en – in het verlengde daarvan – naleving door burgers, bedrijven en instellingen is niet nieuw. Op deze terreinen wordt tot nu toe nog weinig samenhangend wetenschappelijk onderzoek verricht in Nederland en daarbuiten. De Algemene Rekenkamer deed in 2005 onderzoek naar de ‘state of the art’ in de wereld van het toezicht. Een conclusie was dat toezichthouders geen inzicht hebben in de relatie tussen handhaving, naleving en risicoreductie, en daardoor ontbreekt er volgens de AR toch wel wat aan de fundering van hun plannen.

Deze notitie

In deze notitie doen we een voorstel voor een onderzoeksprogrammering EvT. Hoe deze precies tot stand is gekomen, wordt uitgebreid besproken in het onderzoeksrapport dat als bijlage aan dit document is toegevoegd (VOLGT NOG). In die rapportage worden de door ons gemaakte keuzes uitgebreider toegelicht en verantwoord. In deze notitie beperken we ons tot drie onderdelen:

- Om te zorgen dat verschillende onderzoeken in een zekere samenhang kunnen worden uitgevoerd en rekening houdend met onderzoek dat reeds binnen de projecten EvT en HH&G is uitgevoerd, presenteren we een raamwerk. Dit raamwerk biedt ruimte aan multidisciplinair onderzoek en dient als basis voor een afbakening.
- T.b.v. het ondersteunen van rijksinspecties bij de implementatie van effectmeting in de toezichtspraktijk, gaan we in op verschillende onderzoeksdesigns die beschikbaar zijn, geven we een aantal handreikingen en bespreken we thema's die relevant zijn voor inspecties.
- We eindigen met een voorstel voor de *onderzoeksprogrammering* EvT.

Om de relatie tussen EvT en HH&G te benadrukken geven we hieronder een schematische weergave van beide programma's.



2 Definities, raamwerk en afbakening

Toezicht en handhaving zijn inspanningen die er vooral voor moeten zorgen dat regels (beter) worden nageleefd. Er zijn veel alternatieve termen voor deze inspanningen zoals controle, inspectie en opsporing en soms is er sprake van repressief dan wel preventief optreden, hard of zacht toezicht, metatoezicht of systeemtoezicht. Ook voor diegene die regels na moet leven zijn veel termen in omloop. Waar juristen spreken van normadressaten, hebben economen het over agents en spreken sociologen van actoren. In deze paragraaf definiëren we de begrippen toezicht, handhaving en naleving zoals wij die gebruiken in deze notitie. Daarnaast presenteren we in dit hoofdstuk een raamwerk (of analysekader) t.b.v. de programmering EvT.

2.1 Definities

Rechtshandhaving, of kortweg *handhaving*, dient (mede) om de naleving van wet- en regelgeving te bevorderen. Dat wil zeggen, het feitelijke nalevingsniveau (meer) in overeenstemming te brengen met het nalevingsniveau dat de verantwoordelijke politieke bestuurders en handhavers minimaal noodzakelijk en realiseerbaar achten. Handhaving is een samengestelde activiteit, die in engste zin enkel het toepassen van een (civiel-, bestuurs- of strafrechtelijke) sanctie omvat en in ruimste zin, elke handeling die erop is gericht naleving van rechtsregels te bevorderen of overtreding daarvan te beëindigen. Meestal associeert men handhaving met een repressieve manier van optreden tegen regelovertreding, waarbij de repressie bestaat uit het toepassen van dwang. Zo wordt in het rapport 'Landelijke organisaties met handhavende bevoegdheden' (AEF, 2007) onder handhaving verstaan 'elke vorm van daadwerkelijk overheidsoptreden dat is gericht op het voorkomen van regelovertreding of het ongedaan maken van een regelovertreding dat verband houdt met (de bevoegdheid tot) het uitoefenen van dwang'. Als zodanig omvat handhaving zowel bestuursdwang, opsporing en andere vormen van repressief optreden, als toezicht en controle gericht op het ontdekken van regelovertreding. Regelgeving, vergunningverlening en algemene bestuurlijke activiteiten zoals proactief optreden, bestuurlijke preventie en andere vormen van 'verleiding' of 'complicance assistance' die geen enkele relatie hebben met het uitoefenen van dwang vallen niet onder deze formele definitie van handhaving. Dit betekent echter niet dat handhaving los zou moeten staan van deze andere (preventieve of ondersteunende) activiteiten, die op de bevordering van naleving zijn gericht. Zodra het meer in het algemeen gaat om nalevingbeleid waarbij het uitoefenen van dwang een rol kan spelen, wordt echter niet zozeer van handhaving als wel van handhavingsbeleid gesproken. Handhavingsbeleid is breder dan handhaving en heeft betrekking op alle nalevingsbevorderende maatregelen die in verband met het toepassen van dwang worden getroffen. Dat betekent onder meer dat het op orde brengen van de juridische en materiële infrastructuur ook tot het handhavingsbeleid kan worden gerekend (Boek & Visser, 2005). Verder valt bijvoorbeeld ook gedogen niet onder de genoemde formele definitie van handhaving, terwijl dit wél deel kan uitmaken van handhavingsbeleid. In feite gaat het bij handhavingsbeleid om een ruimere, meer materiële definitie van wat onder handhaving wordt verstaan. Wij kiezen voor deze "ruimere" opvatting.

Onder *toezicht* wordt verstaan het verzamelen van informatie over de vraag of organisaties, beleid of voorzieningen aan de gestelde eisen voldoen, het zich daarna vormen van een oordeel daarover en het eventueel naar aanleiding daarvan interveniëren (Commissie Holtslag, 1998). Op deze wijze opgevat is toezicht een smaller begrip dan handhaving, omdat het zich specifiek op de beoordeling en interventie van concrete gedragingen richt. Organisaties die toezicht uitoefenen, heten doorgaans inspectie (Inspectie Werk en Inkomen), Autoriteit (Autoriteit Financiële Markten) of Commissariaat (van de Media). Onderscheid wordt wel gemaakt tussen kwaliteitsmeesters (zoals de Onderwijsinspectie), marktmeesters (zoals de Opta en de NMA) en regelmeesters (zoals de Arbeidsinspectie). Ook verschillende objecten van toezicht worden (o.a. door de Ambtelijke Commissie Toezicht) onderverdeeld in toezicht op zelfstandige organisaties en professionals met een publieke taak, toezicht op een

publieke of een beleidssector, toezicht op marktsectoren en toezicht op de naleving van aspectwet- en regelgeving. Meestal echter worden nalevingstoezicht, uitvoeringstoezicht, interbestuurlijk en markttoezicht als toezichtsvormen onderscheiden. Als toezichthouders dwang uitoefenen, wordt van *handhaving* gesproken (Commissie Kuipers, 2005). De essentie van toezicht is dat een actor kan ingrijpen in (corrigerend kan optreden bij) de besluiten of handelingen van een verzameling andere actoren. De essentie van handhaving is dat een actor andere actoren uit een gedefinieerde verzameling (normadressaten) ertoe kan bewegen zich overeenkomstig bepaalde regels te gedragen. Toezicht is, in dat licht bezien, juist een ruimer begrip. Toezicht omvat een grotere verzameling betrekkingen tussen actoren dan handhaving (De Ridder, 1988). Een belangrijke taak van een toezichthouder is te zorgen dat wetten en regels worden nageleefd door burgers, bedrijven en de overheid zelf. Controleren van naleving is echter geen op zichzelf staand doel maar dient bij te dragen aan het voorkomen en beheersen van collectieve risico's die bepaalde gedragingen van onder toezichtgestelden met zich meebrengen en dient (positieve) welvaartseffecten te helpen genereren zoals betere gezondheid, betere opleidingen, veiliger verkeer en vervoer enz.

Onder *regelnaleving* (of compliance) wordt verstaan het zich feitelijk gedragen in overeenstemming met bepaalde regels, waaronder ook begrepen wordt het zich onthouden van gedrag dat door die regels wordt verboden (Elffers en Hessing, 1992). Naleving is een neutraal begrip: elke afwijking van de regels wordt als niet-naleving gezien, of deze afwijking nou opzettelijk of onopzettelijk plaatsvindt, en of naleving nou vrijwillig is of wordt afgedwongen door strafdreiging. Naleving wordt in de Angelsaksische literatuur compliance genoemd en deze term wordt in de Nederlandse literatuur ook regelmatig gebruikt. Tegenover naleving staat het niet naleven van de regels. Niet-naleving kan ook als overtreding worden aangeduid, alleen kleeft aan deze term de associatie van opzettelijkheid. Naleving zou in enge zin opgevat kunnen worden als het naleven van wettelijke regels. Er zijn echter heel veel andere regels die nageleefd kunnen of moeten worden, alleen zijn die niet vastgelegd in wetten maar bijvoorbeeld in private reguleringsarrangementen zoals convenanten, verdragen of overeenkomsten. Dan zijn er tenslotte ook nog informele en sociale regels die nageleefd of overtreden kunnen worden.

2.2 Raamwerk EvT en HH&G

In algemene zin kan 'effectiviteit van handhaving en toezicht' omschreven worden als de mate waarin de uitoefening hiervan leidt tot verwezenlijking van de geformuleerde doeleinden. Veel onderzoek beoogt de werking vast te stellen van een bepaald toezichtsoptreden op een bepaald moment in termen van een bepaalde gedragsverandering. In het onderzoeksprogramma EvT focussen we op twee effecten: *nalevingseffecten* en *maatschappelijke effecten*. Toezichthouders kunnen zich richten op het versterken van naleving, maar ook op het behalen van maatschappelijke doelen. Hoe deze effecten (op individueel niveau) tot stand komen, wordt uitgebreid onderzocht binnen het programma HH&G. Daar is de centrale vraag hoe het nalevingsgedrag van burgers, bedrijven en instellingen wordt beïnvloed door (1) individuele motieven en kenmerken, (2) door de wijze van toezicht en handhaving of door (3) andere omgevingsfactoren. In het raamwerk dat we in deze paragraaf presenteren, werken we deze invloeden uit en geven we deze een plek.

Bij de ontwikkeling van het raamwerk hebben wij gebruik gemaakt van het zogenaamde bootje van Coleman³. De belangrijkste idee achter dit model is dat voor de verklaring van gebeurtenissen op macroniveau, gekeken moet worden hoe relevante kenmerken van de omgeving of gebeurtenissen in die omgeving, het gedrag van individuele actoren beïnvloeden en hoe dit gedrag leidt tot nieuwe macrosituaties. Met dit model wilde Coleman verduidelijken hoe het gedrag van sociale systemen zich verhoudt tot het

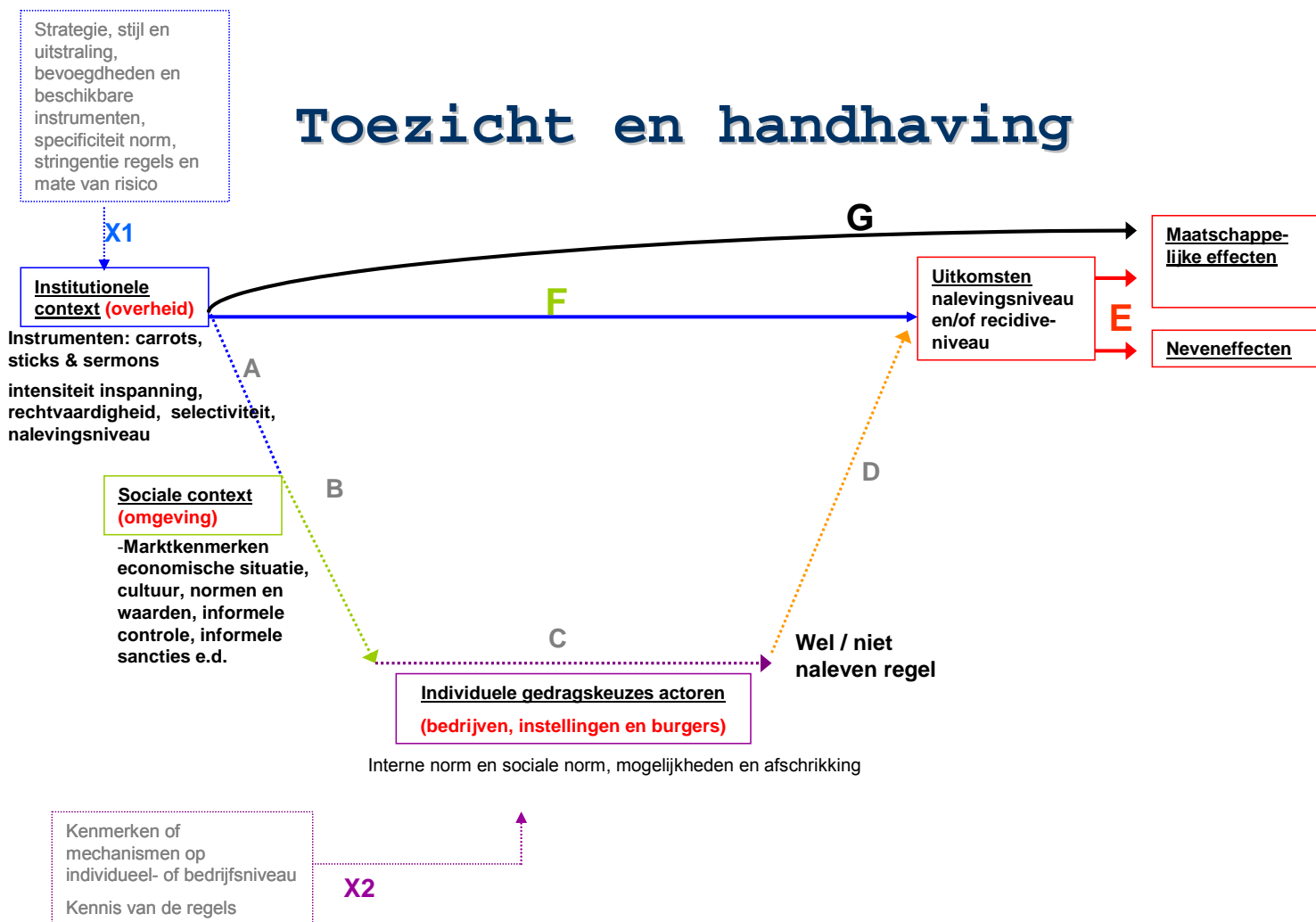
³ In het onderzoeksrapport dat aan deze notitie ten grondslag ligt, gaan wij uitgebreid in op het bootje van Coleman.

gedrag van de individuen waaruit deze sociale systemen zijn opgebouwd en hoe moet worden gedacht over het aggregeren van het individuele microgedrag tot macro-uitkomsten. Vanuit deze gedachte is een analysemodel van gedrag en gedragsregulering ontwikkeld dat gedragsmechanismen een plaats geeft met het oog op beleidsinterventies meer houvast kan beiden waar het om het onderscheiden van oorzaken, condities en gevolgen gaat. Het model combineert drie soorten theorieën:

- Microtheorie over de acties of keuzes van individuen. Gedragsmechanismen zijn het best te begrijpen als algemene verklaringen voor de keuzen die individuele actoren maken of de gedragingen die zij vertonen (C).
- Macro-microtheorie die aangeeft hoe de keuzen van actoren of de gedragingen die zij vertonen worden beïnvloedt door de mogelijkheden of beperkingen die de omgeving waarin zij zich bevinden biedt (A en B).
- Micro-macrotheorie die aangeeft hoe de verschillende keuzen of gedragingen van individuele actoren optellen tot uitkomsten op het niveau van de samenleving (D).

Substitueren we toezicht-, handhaving- en nalevingsaspecten in het model, dan ontstaat het raamwerk. Figuur 1 geeft onze poging weer om een generiek, overkoepelend kader te ontwikkelen, waarbinnen verschillende perspectieven en theoretische stromingen een plek vinden. Het raamwerk biedt tevens mogelijkheid om het project EvT en HH&G af te bakenen.

Figuur 1: Raamwerk project EvT en HH&G



Het raamwerk stelt het gedrag van actoren centraal (C). Vervolgens laat het raamwerk zien dat dit gedrag wordt beïnvloed door een complex van omgevingsinvloeden (A en B). Hoe het nalevingsgedrag van actoren kan worden verklaard vanuit de kenmerken en motieven van de doelgroep en vanuit de omgevingsfactoren, is hiermee afgedekt. Een vervolgvraag is hoe individueel gedrag optelt tot een bepaald nalevingsniveau (D) en hoe dit niveau uiteindelijk bijdraagt aan het bewerkstelligen van maatschappelijke effecten (E). Tenslotte onderscheiden we F en G en daarmee bedoelen we de causale relatie tussen interventies en resp. nalevingsniveau en maatschappelijk effect op macroniveau.

Leggen we de doelstelling van het project EvT aan tegen het raamwerk, dan kunnen we een *afbakening* formuleren. In het programma EvT zijn drie niveaus te onderscheiden die ook als niveau terugkomen in het raamwerk; EvT moet gaan over de mate waarin het toezicht- en handavingsactiviteiten van invloed zijn (1) op nalevings- en recidiveniveau, (2) op maatschappelijke uitkomsten en (3) hoe (meer) naleving of (minder) recidive zorgt voor maatschappelijke effecten. We hebben het dan over de onderdelen **E, F, G en X1** uit het raamwerk. De overige onderdelen (A t/m D en X2) komen aan uitgebreid bod binnen het onderzoeksprogramma HH&G.

Het effect van toezicht- en handavingsinstrumenten op nalevings- of recidiveniveau (F) en op maatschappelijke uitkomsten (G)

Met *institutionele context* bedoelen we in figuur 1 de wijze waarop de overheid in het algemeen maar toezichthouders en handhavers in het bijzonder, met behulp van hun instrumentarium, invloed uit kunnen oefenen op de kennis, motieven en mogelijkheden van actoren. Een algemene indeling van beleidsinstrumenten bespreken we in het onderzoeksrapport (carrots, sticks and sermons). Om de werking het instrumentarium en de invloed op gedrag goed te bestuderen, is het nodig verder te kijken dan naar de werkzaamheid van toezicht en handhaving alleen. De regelgeving die er vaak aan ten grondslag ligt dient meegenomen te worden. De keuze van het instrumentarium is weer afhankelijk van allerlei factoren (X1). Voor handhavende of toezichthoudende organisaties geldt o.a. dat de mate van risico dat beperkt dient te worden, de beschikbaarheid van instrumenten en de bevoegdheid om deze te mogen gebruiken, strategie en stijl, de stringentie van de onderliggende regel en de specificiteit van de norm, meespelen in de instrumentkeuze en in de wijze waarop deze worden ingezet. Institutionele en sociale context worden in het model onderscheiden t.b.v. de ordening van studies in het vervolg van dit rapport. In werkelijkheid bestaat dit onderscheid niet echt en is er sprake van interactie tussen factoren die tot beide contexten behoren. Soms werken ze samen maar ze kunnen ook concurreren met elkaar. Het publiceren van inspectieresultaten bijvoorbeeld, zal een groter effect hebben binnen sectoren waar reputatie een belangrijke rol speelt in het verkrijgen van opdrachten, dan binnen sectoren waar reputatie of resultaten uit het verleden geen enkele rol spelen. Het is met andere woorden noodzakelijk bij de keuze van een instrument (ook) rekening te houden met concurrerende mechanismen die door de sociale context geactiveerd worden.

Hoe naleving zorgt voor maatschappelijke- en neveneffecten (E)

Onder optimale omstandigheden zorgt een hoog nalevingsniveau voor maatschappelijke effecten. Denk bijvoorbeeld aan de reductie van het aantal verkeersdoden, minder stomdronken pubers of gemiddeld hogere Citoscores van groep-achters. Maatschappelijke effecten worden niet altijd bereikt en we noemen drie oorzaken. Allereerst kan er sprake zijn van *theory failure*; de aan de aanpak ten grondslag liggende beleidstheorie is onjuist. De veronderstelde mechanismen werken bijvoorbeeld niet of op een andere manier dan verwacht en de opbrengsten blijven daardoor uit. Maar het kan ook zijn dat interventies perverse of onbedoelde neveneffecten teweeg brengen waarmee in de ontwerpfasen geen rekening is gehouden. Een tweede oorzaak is *implementation failure*. Daarvan is bijvoorbeeld sprake als er (om pragmatische redenen) bij de implementatie wordt afgeweken van het plan. Dergelijke fouten leiden ertoe dat de veronderstelde mechanismen niet in werking kunnen treden en het effect daardoor uitblijft (of zelfs tegengesteld is aan wat werd beoogd). *Measurement failure* is de derde

oorzaak die we noemen. De beleidstheorie kan in orde zijn, de interventie kan goed zijn geïmplementeerd, maar als bij het meten van de effecten fouten worden gemaakt, lijkt het alsof de aanpak toch 'mislukt' is. Dit kan gebeuren als de geraadpleegde informatiebronnen bijvoorbeeld niet betrouwbaar of niet valide zijn, het onderzoeksdesign niet deugt of onderzoekers hun werk niet goed hebben gedaan. Of een hoger nalevingsniveau van een regel ook daadwerkelijk zorgt voor maatschappelijke effecten, hangt dus van een aantal zaken af. Als het plan klopt, de implementatie juist is uitgevoerd en de werkzame elementen worden geactiveerd en valide en betrouwbaar onderzoek laat zien dat het nalevingsniveau significant hoger is (door toedoen van de interventie) en dit heeft geleid tot het einddoel, dan kan men waarlijk spreken van maatschappelijke effecten.

3 EvT: Onderzoeksdiseins, handreikingen en thema's

3.1 Designs

Bij de vaststelling van effecten van toezicht gaat om het vaststellen van een causaal verband tussen gedragingen van onder toezichtgestelden 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 toezichtgestelden, 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 (Winter & de Ridder, 2010).

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 toezichtgestelden (normadressaten) bij wie het te onderzoeken optreden van de toezichthouder niet plaats vindt. Ook bij deze controlegroep moeten een voormeting en een nameting of een meting in een tijdreeks uitwijzen of andere factoren in het spel zijn. Kan bij de controlegroep geen verandering worden vastgesteld dan is de conclusie dat het toezicht het veronderstelde effect had op zijn minst aannemelijk.

Effectiviteitsonderzoek eist onderzoeksdiseins 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 onderzoeksdiseins 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 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.

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).

Het is raadzaam om een zo hoog mogelijke MSMS-score na te streven maar de stelregel is dat vanaf niveau 3, men kan spreken voldoende interne validiteit. Vanaf niveau 3 is het verantwoord om waargenomen verschillen in een uitkomstvariabele (outcome) toe te schrijven aan de ingezette interventie of maatregel.

3.2 Handreikingen effectonderzoek

Winter & de Ridder (2010) vonden 70, vooral buitenlandse effectstudies en zij beschreven deze in termen van effectiviteit, gebruikte onderzoeksdesign en methoden en technieken. Zij concludeerden dat effectonderzoek niet makkelijk is maar zeker wel mogelijk en op basis van hun analyses formuleerden zij een aantal handreikingen.

1. *Eenvoud*

- Het meten van nalevingsniveau is eenvoudiger dan het meten van maatschappelijke effecten. Het is verstandig om ervaring op te doen met het meten van effecten van toezicht op naleving en dit daarna uit te breiden met maatschappelijke effecten.
- Naast het meten van maatschappelijke effecten is het ook raadzaam na te gaan of er bestaande databestanden voorhanden zijn waarin deze kenmerken zijn vastgelegd. Alhoewel er wel wat haken en ogen aan het gebruik van bestaande registraties zitten (denk aan de mate van betrouwbaarheid en een mogelijke bias), is het gebruik ervan efficiënt.
- Ervaring met effectonderzoek dient stukje bij beetje te worden opgebouwd, beginnend met eenvoudige designs, waarbij wordt getracht een enkelvoudige toezichtinterventie te relateren aan een specifieke groep normadressaten.

2. *Reconstrueer de beleids- of programmatheorie*

- Goed effectiviteitsonderzoek begint met een analyse van de bestaande toezichtsstrategie. Het is belangrijk om te expliciteren hoe het gedrag van de normadressaat wordt beïnvloed door de ingezette interventie en hoe dit uiteindelijk resulteert in het beoogde maatschappelijke effect. Bij het ontwikkelen van een onderzoeksdesign is het goed te weten hoe de beleids- of interventietheorie in elkaar steekt. Soms bestaat deze al maar meestal moet deze nog ontwikkeld worden. Door oorzaak-gevolg relaties uit te werken wordt vooraf duidelijk hoe effecten tot stand zouden moeten komen. Vervolgens dient het effectonderzoek uit te wijzen of die causale relaties ook werkelijk tot stand zijn gekomen.
- Door een goede reconstructie komen ook positieve en negatieve neveneffecten boven tafel. Onderzoekers moeten zich bewust zijn van mogelijke "moral hazards". Door voorafgaand aan het effectonderzoek op de hoogte te zijn van mogelijke neveneffecten, kan hier rekening mee gehouden worden.

3. *Doe ook kwalitatief onderzoek*

- Niet alleen kwantitatief onderzoek onder grote aantallen normadressaten naar de aanwezigheid van causale relaties tussen een interventie en de effecten daarvan is waardevol. Verklarend (effectiviteits)onderzoek is ook mogelijk met behulp van kwalitatieve methoden van dataverzameling. Interviews en observaties kunnen de resultaten van kwantitatief effectonderzoek enorm verrijken.

4. *Let op de samenstelling van de controlegroep*

- Bij het uitvoeren van MSMS3+ onderzoek worden normadressaten idealiter at random ingedeeld in de interventie- en de controlegroep. Dit is niet altijd mogelijk en dan zijn er andere mogelijkheden om bias in de resultaten te voorkomen. Een mogelijkheid is gebruik te maken van het "matching pair" principe. Als er geen

sprake is van random toewijzing dan is het altijd verstandig om in de analysefase te controleren voor andere gemeten kenmerken die voorhanden zijn.

5. *Inbedding van effectonderzoek in de toezichtsstrategie*

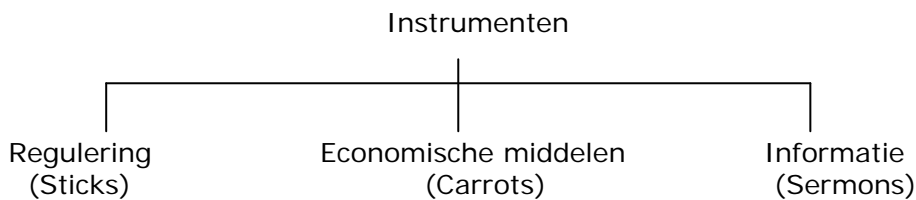
- Effectonderzoek moet zoveel als mogelijk is realistische toezichtssituaties evalueren. Ook experimenten dienen te zijn ingebed in een bestaande toezichtsstrategie.

3.3 Thema's

Wie de literatuur bijhoudt over handhaving, toezicht en naleving binnen verschillende wetenschappelijke disciplines, beleidsnota's leest of de berichtgeving hierover volgt in de media, kan al snel een arsenaal aan thema's bedenken dat het onderzoeken waard is. In dit hoofdstuk zullen we er een aantal noemen die passen binnen de afbakening EVT (zie figuur 1: E, F en G). We gebruiken de thema's om de voorgenomen studies verder in te richten en om te zorgen voor herkenbaarheid bij de participanten EvT.

Beleidsinstrumenten

Toezicht houden is gedrag beïnvloeden. Voor die beïnvloeding hebben toezichthouders en (meer algemeen gesproken) de overheid nogal wat instrumenten in huis. Vedung e.a. (1998) geven een overzicht van typologieën waarmee overheidsinterventies in het algemeen geordend kunnen worden. Via een aantal tussenstappen komen zij tot de volgende:



De instrumenten verschillen in de mate van vrijheid die zij de normadressaat laten om iets te doen of om iets op haar eigen manier te doen. Regulering is het meest bindende beleidsinstrument, terwijl informatieverstrekking de geadresseerde de meeste vrijheid van handelen biedt. In het geval van regulering is de normadressaat verplicht te doen wat de wetgever hem opdraagt te doen. In het geval van economische middelen is de normadressaat niets verplicht maar wordt het bepaald gedrag wel gestimuleerd of juist afgeremd met behulp van materiële middelen terwijl het bij informatieverstrekking gaat om overreding met behulp van communicatiemiddelen.

Overheidsinstrumenten worden zelden separaat ingezet, meestal komen ze in stapels of pakketten. Er zijn drie soorten pakketten die we verticale, horizontale en chronologische stapels noemen. Met verticale stapeling wordt bedoeld dat de overheid meerdere bestuurslagen gebruikt om uiteindelijk een doelgroep te bereiken. Met behulp van instrumenten oefenen hiërarchisch hogere organen invloed uit op intermediairs. Dit zijn instanties lager in de hiërarchie (zoals toezichthouders of agentschappen), die op hun beurt weer instrumenten inzetten die een doelgroep moeten bereiken. Het kan gaan om één of meer intermediairs voordat de uiteindelijke doelgroep wordt bereikt. Intermediairs hoeven overigens geen deel uit te maken van de overheid. De overheid kan bijvoorbeeld wettelijk vastleggen dat voedselproducenten aan moeten geven hoeveel calorieën een product bevat (regulering). Door dit te verplichten zijn het de producenten die de consumenten voorlichten over de (on)schadelijkheid van een product (informatie verstrekken).

Horizontale stapeling betekent dat twee of meer instrumenten tegelijkertijd worden ingezet door dezelfde organisatie. Responsive regulation en andere zogenaamde give and take strategieën zijn daar voorbeelden van. De centrale idee is dat eerst wordt getracht met zachte hand, actoren te beïnvloeden en als dat niet lukt over te stappen op meer

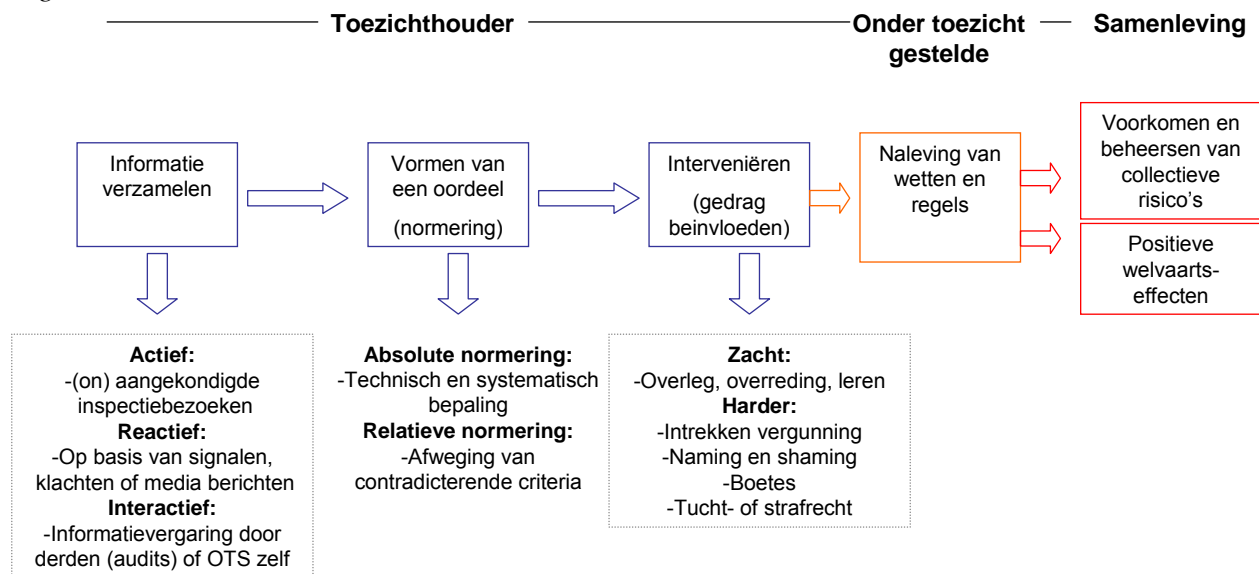
strikte instrumenten. Binnen een doelgroep kan dit betekenen dat de goedwillende meerderheid advies krijgt over bepaald ongewenst gedrag terwijl een hardnekkige, bewuste overtredende minderheid een sanctie opgelegd krijgen voor soortgelijk gedrag. Als er sprake is van een chronologische inzet van instrumenten, dan wordt van chronologische stapeling gesproken. De 3-E's strategie (education, engineering and enforcement) is hier een voorbeeld van. Een bepaald probleem wordt dan door de tijd heen aangepakt door eerst informatie te verstrekken om bij onvoldoende resultaat over te stappen op de inzet van 'incentives' en pas op het laatst te kiezen voor 'rules and regulation'.

Alhoewel deze principes ook gelden voor toezichthouders en inspecties, zijn er binnen het toezichtsdomain wel een aantal specifieke stappen te onderscheiden. Eerder (par. 2.1) gaven we aan dat toezicht grofweg kan worden ontleed in een aantal componenten:

- informatie over de vraag of organisaties, beleid of voorzieningen aan de gestelde eisen voldoen;
- het zich daarna vormen van een oordeel daarover;
- en het eventueel naar aanleiding daarvan interveniëren.

Daar kwam bij dat een belangrijk doel van een toezichthouder is te zorgen dat wetten en regels worden nageleefd door burgers, bedrijven en de overheid zelf. Uiteindelijk dient dit bij te dragen aan het voorkomen en beheersen van collectieve risico's die bepaalde gedragingen van onder toezichtgestelden met zich meebrengen en dient (positieve) welvaartseffecten te helpen genereren. Geven we de logica achter toezicht sterk vereenvoudigd weer, dan ontstaat de volgende figuur.

Figuur 2.



Als we er vanuit gaan dat de bovenstaande beleidslogica correct is dan kunnen we stellen dat de toezichtscomponenten uit figuur 2 direct of indirect van invloed zijn op de mate van naleving en op de maatschappelijk effecten. Daarmee passen ze binnen de afbakening en zijn ze relevant voor het programma EvT.

Informatie over het gedrag van normadressaten kan actief worden vergaard wanneer toezichthouders zelf op inspectie gaan. Vooral de klassieke inspecties maken gebruik van een dergelijke vorm van actieve informatievergaring. Toezichthouders kunnen ook reactief informatie verzamelen, bijvoorbeeld wanneer er signalen worden ontvangen waaruit blijkt dat er mogelijk iets mis is. Interactieve informatievergaring vindt plaats wanneer toezichthouders aan anderen opdragen om periodieke rapportages aan te leveren waarover de toezichthouder vervolgens haar oordeel vormt. Dat kunnen

algemene documenten zijn zoals jaarverslagen, maar ook zogenaamde incident- of calamiteitenmeldingen. Een zinvolle vraag die gesteld kan worden is wat het effect van de verschillende vormen van feitenverzameling is op nalevingsniveau en maatschappelijke uitkomsten.

Dezelfde vraag kan gesteld worden over de component *normering en oordeelsvorming*. Deze kunnen eveneens op verschillende wijzen plaats vinden. In sommige gevallen is de norm het resultaat van technocratische en systematische analyses van gelijke of vergelijkbare gevallen. In andere gevallen is de norm eerder het resultaat van onderhandelingen met betrokken partijen. Normen kunnen een absoluut karakter hebben, denk bijvoorbeeld aan minimale solvabiliteitscriteria voor een financiële instelling, maar kunnen ook het resultaat zijn van een afweging van verschillende contraderende criteria. Te denken valt aan het gebruik van behandelrichtlijnen in de medisch curatieve gezondheidszorg waarin een afweging wordt gemaakt tussen medisch noodzakelijke en effectieve interventies enerzijds en de kosten die aan die interventies verbonden zijn anderzijds. Uiteindelijk dient normering uit te monden in een toetsingskader dat door alle partijen wordt gekend. De wijze waarop normering vorm krijgt is van invloed zijn op de oordeelsvorming door een toezichthouder, op de wijze van interveniëren en uiteindelijk op naleving en maatschappelijk effect.

Interveniëren (om gedrag te beïnvloeden) kan op verschillende manieren. Voor toezichthouders geldt ook dat ze beschikken over verschillende soorten instrumenten die regelmatig gestapeld worden. In een compliance-achtige setting passen de meer zachte instrumenten van overreding, overleg, diplomatie of onderricht. Toezichthouders moeten dan wel het vertrouwen hebben dat bedrijven of burgers de adviezen omzetten in gewenst gedrag. In een deterrence-achtige setting passen de hardere instrumenten van sancties als boetes, het intrekken van vergunningen of uitsluiting. Toezichthouders zijn in de afgelopen 10-15 jaar komen te beschikken over meer en meer gedifferentieerde sancties. Naast waarschuwingen en het intrekken van vergunningen is in circa 60 wetten voorzien dat bestuurlijke boetes opgelegd kunnen worden (Michiels, 2006) die soms kunnen oplopen tot vele miljoenen euro's. Ook worden reparatoire sancties toegepast, zoals de 'last onder dwangsom' waarmee de toezichthouder beoogt een halt toe te roepen aan een overtreding en erop uit is om, waar mogelijk, reparatie te laten plaatsvinden of de schade te laten herstellen. Bestuursdwang functioneert als stok achter de deur of voor gevallen waarin geen overtreders bekend zijn. Daarnaast is er de strafbeschikking van de politie en de OM-afdoening (de strafbeschikking in handen van het OM) (vgl. Elsink, 2007, 20-22). Tenslotte wordt met naming and shaming gewerkt, waarbij personen en organisaties met naam en toenaam aan de schandpaal genageld worden.

Huisman (2007) liet zien dat de bestuurlijke boete de meest toegepaste sanctie tegen bedrijven is en dat op tal van beleidsterreinen de bestuurlijke boete de strafrechtelijke sanctionering heeft verdrongen. De reden hiervan was weliswaar de tekortschietende strafrechtelijke handhaving jegens bedrijven, maar de bestuurlijke boete mist het morele stigma dat is verbonden aan een door de strafrechter opgelegde sanctie. Onderzoek naar het effect van verschillende soorten sancties is (met de constatering van sanctierijkdom) daarom belangrijk.

Stijlen, strategieën en risico's

De wijze van informatieverzameling door toezichthouders, de oordeelsvorming naar aanleiding hiervan en het interveniëren, kan op verschillende manieren. De keuzes die worden gemaakt door de toezichthouder bepaalt de stijl van toezicht. Deze wordt niet alleen beïnvloed door de ernst van het risico dat moet worden beperkt, de mate van (spontane) naleving van regels en wetten of het karakter van de onder toezichtgestelden, maar ook door de publieke en politieke discussie. In deze discussie doet zich nu en in de afgelopen jaren een paradoxaal verschijnsel voor. Enerzijds is er de tendens van politiek en samenleving om na incidenten en crises op te roepen tot meer toezicht (en regels). De discussies over de rol van de Nederlandse Bank en de Autoriteit Financiële Markten in het Icesafe debacle is daar een recent voorbeeld van. In de Brief

van de Minister van BZK over het programma 'Vernieuwing Toezicht' (ruim voor de kredietcrisis) heette het zo: 'na incidenten vragen politiek en samenleving doorgaans om versterking van het toezicht'. Ook het adagium 'Pech moet weg' (Mertens et al, 2002) figureert ook in de maatschappelijke discussie en maakt duidelijk dat de afgelopen jaren van toezichthouders en handhavers steeds meer verwacht wordt dat zij risico's indammen, zo niet voorkomen. De roep om meer wetten, regels, toezicht en handhaving ligt vanuit dit concept voor de hand, zeker na ernstige incidenten. Anderzijds roepen burgers en onder toezicht gestelde organisaties tegelijkertijd dat ze te veel last hebben van bemoeienissen van toezichthouders, door de 'druk(te)' die ze veroorzaken, de administratieve lasten die daarmee gepaard gaan en omdat lang niet altijd duidelijk is hoe toezicht bijdraagt aan het gemeenschappelijk belang. Toezicht moet kortom anders, efficiënter en effectiever en hieraan gerelateerd gaat het Kabinet er van uit dat inspecties zoveel mogelijk gebruik gaan maken van controlesystemen die in de onder toezicht staande organisaties en ketens zelf al bestaan. (vgl. de Brief van de Minister van BZK over het programma 'Vernieuwing Toezicht' (Min BZK, 31 201 25)). Er wordt met andere woorden gepleit voor (en nogal wat verwacht van) trust-and-hope regulation, niet alleen in Nederland maar ook in andere westerse landen (Shover, 2008). De orthogonaal van "trust-and-hope regulation" is formalistisch toezicht dat ook wel als "command and control" of "going by the book" wordt getypeerd (Erp, 2008). Deze vorm van toezicht is gebaseerd op wantrouwen en afschrikking en leunt sterk op bestraffing. Uiteraard vinden vertrouwen en wantrouwen en trust-and-hope en command and control elkaar in een mix. Responsive regulation (Braithwaite, 1992) is de meest bekende strategie die gebruik maakt van beide uitersten. Responsive regulation gaat uit van de handhavingspiramide die aansluit bij de gedachte "niet goedschiks dan maar kwaadschiks" en is in de afgelopen jaren uitgegroeid tot de dominante strategie van Nederlandse handhavingsorganisaties (Van de Bunt, 2007). Huisman (2007) merkt op dat beleidsmakers zeer gretig zijn als het gaat om het kiezen van nieuwe strategieën. Voorbeelden betreffen de handhavingspiramide van Braithwaite e.a., het stimuleren van zelfregulering en certificering, deregulering en marktwerking en het idee dat zwaarder straffen de bedrijfsinterne besluitvorming beïnvloedt. Voor sommige van deze theorieën en modellen geldt overigens wel dat ze zijn geformuleerd op basis van de resultaten van empirisch onderzoek, zoals het model van Braithwaite, maar ze zijn vervolgens zelf onvoldoende empirisch getoetst. Hoewel sommige modellen mede dankzij deze empirische grondslag bijzonder overtuigend lijken, kan de waarde van deze modellen worden verhoogd wanneer ze op basis van de resultaten van empirisch onderzoek verder worden bijgewerkt en ontwikkeld. Het is zinvol om meer empirisch onderzoek te doen naar de effecten van strategieën van regulering, toezicht en handhaving op het gedrag van bedrijven instellingen en burgers.

De combinatie van een roep om beter toezicht en de erkenning dat dit slechts onder bepaalde voorwaarden mogelijk is, vormt de kern van verbeterpogingen die rijks- en andere overheden in de afgelopen jaren hebben ondernomen. Daarbij gaat het steeds meer om het terugbrengen van risico's voor burgers en steeds minder om toezicht op rechtmatigheid, doelmatigheid en doeltreffendheid. Het credo 'Pech moet weg' heeft er o.a. toe geleid dat er steeds minder wordt opgetreden op basis van geconstateerde overtredingen, maar op basis van de inschatting van risico's. De hiertoe ontwikkelde modellen zijn in belangrijke mate geïnspireerd door methoden van risk assessment in het bedrijfsleven (Huisman, 2007). Een belangrijke vraag is hoe betrouwbaar en valide deze risicotaxaties zijn en in welke mate deze in staat zijn de grootste maatschappelijke risico's te neutraliseren.

Samengevat moet het programma EvT antwoord geven op de vraag of het al dan niet inzetten van de genoemde sanctiesoorten, gegeven de wijze van informatieverzameling, normering en oordeelsvorming, zorgt voor meer naleving, meer maatschappelijk effect en minder pech. Daartoe behoren ook de stijlvormen en de vormen van risico-inschatting die inspecties hanteren. Dan is er tenslotte nog de economische dimensie: toezicht mag niet te veel kosten en je mag er ook niet al te veel last van hebben.

4 Voorstel programmering EvT

In dit hoofdstuk doen we een voorstel voor het onderzoeksprogramma Effecten van Toezicht. Bij de samenstelling ervan hebben we ons steeds afgevraagd of een drietal vragen bevestigend konden beantwoorden:

1. Past het voorgestelde onderzoek binnen de afbakening van het raamwerk (par. 2.2)?
2. Gebruiken we de gepresenteerde designs en handreikingen (par. 3.1 en 3.2)?
3. Is het onderzoek relevant binnen de politieke en maatschappelijke discussie over toezicht (par. 3.3)?

Binnen de programmering maken we een onderscheid tussen experimentele van niet-experimentele onderzoeken. Bij de experimentele studies ligt de nadruk op het vaststellen van causale relaties tussen nalevingsniveau of maatschappelijk effect enerzijds en toezichtsinterventies anderzijds. Door te kiezen voor experimentele designs is het (beter) mogelijk oorzaak-gevolg relaties vast te stellen. Voor deze onderzoeken is participatie van de inspecties een vereiste. Externe onderzoekers zullen helpen bij het ontwikkelen van het juiste design en het berekenen en interpreteren van de resultaten. De studies die we hieronder beschrijven, zijn voor wat betreft het onderwerp aan te passen, het belangrijkste doel is het stimuleren van intern valide effectonderzoek (MSMS 3+).

Bij de niet-experimentele studies ligt de nadruk vooral op het inventariseren van voor dit programma relevante kennis, bijvoorbeeld de beschikbaarheid en betrouwbaarheid van externe bronnen of de mogelijkheden van het structureel toepassen van random audits.

Na vaststelling van de programmering door de projectgroep EvT en de Inspectieraad, zullen de onderzoeksvoorstellen verwerkt worden tot startnotities. Doel- en probleemstelling, onderzoeksvragen, achtergronden en mogelijke methoden en technieken komen dan meer uitgebreid aan de orde. De startnotities dienen vervolgens als startpunt voor het aanvragen van offertes bij mogelijke uitvoerders en als leidraad voor gecontracteerde onderzoekers en de begeleidingscommissies. Een begeleidingscommissie t.b.v. een onderzoek uit EvT, heeft altijd een onafhankelijke voorzitter en daarnaast wordt een van de leden van de werkgroep EvT gevraagd plaats te nemen.

4.1 Experimentele studies

Studie 1

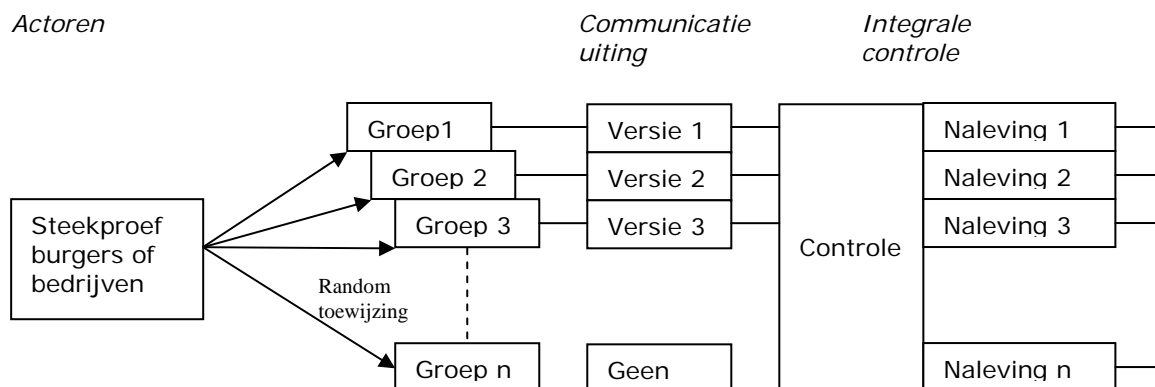
Titel: Invloeden van handhavingscommunicatie op naleving van regels
Raamwerk: F
Type: Veldexperiment

Korte beschrijving:

De literatuurstudie van Van Erp (2007) laat zien dat over de effecten van communicatie-inspanningen op de naleving van wettelijke regels door burgers of bedrijven nog maar weinig bekend is. In het algemeen kan worden gesteld dat handhavingsinspanningen effectiever zullen zijn als zij ingrijpen op de dominante factoren achter regelnalevend of regelovertredend gedrag en in hoofdstuk 2 noemden we er verschillende.

Inmiddels is er een senario-achtige studie gestart naar het effect van verschillende communicatiestijlen op het voorgenomen nalevingsgedrag van personen en bedrijven. Deze studie wordt in de loop van 2010 afgerond en de uitkomsten ervan kunnen als input dienen voor de studie die we hier voorstellen (studie 1). In deze studie willen we de resultaten vertalen naar een echte toezicht- of handhavingssituatie. Denk bijvoorbeeld aan het bekendmaken van een nieuwe regel m.b.t. de opslag van vuurwerk, het verbod op het gebruik van een milieubelastend middel of het juist invullen van een aangifte. De

idee is om de belangrijkste determinanten die uit de scenario-studie naar voren komen, te verwerken in brieven of andersoortige communicatie-uitingen en deze te versturen aan echte actoren. De uitingen komen in verschillende versies waarbij wordt gevarieerd in determinanten en de boodschap. Uit een bestand met onder toezichtgestelden (bedrijven of personen) wordt een steekproef getrokken en deze wordt at random verdeeld over de verschillende communicatievarianten. Eén groep ontvangt geen enkele uiting (controlegroep). Vervolgens worden alle geadresseerden gecontroleerd en wordt er bepaald of de in de brief genoemde regel ook daadwerkelijk wordt nageleefd. Geven we het design schematisch weer dan ontstaat de volgende figuur:



Wat het oplevert:

De uitkomsten van studie 1 kunnen worden gebruikt voor de ontwikkeling van meer effectieve communicatiestrategieën. Door systematisch te variëren in relevante communicatiekenmerken en vervolgens te controleren in hoeverre normadressaten zich aan de gecommuniceerde regel houden, ontstaat inzicht in het relatieve belang van deze kenmerken. Wellicht is het in bepaalde omstandigheden of in bepaalde contexten verstandiger om neutraal te communiceren met onder toezichtgestelden terwijl in andere gevallen een strenge toon juist zorgt voor meer naleving.

Studie 2

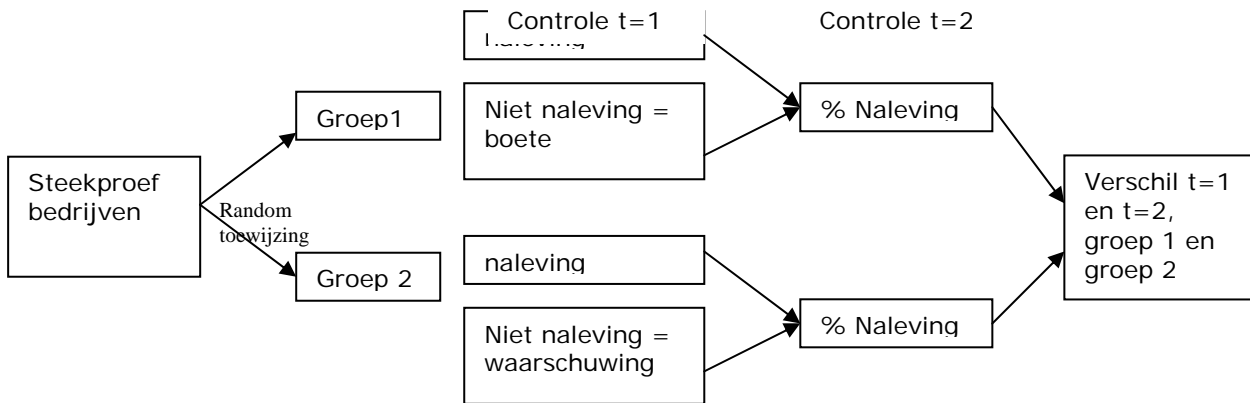
Titel: De effecten van bestuurlijke sancties op naleving
 Raamwerk: F
 Type: Veldexperiment

Korte beschrijving:

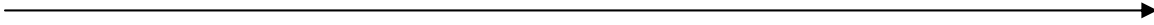
Huisman e.a. gaven aan dat het bij het bepalen van de straf essentieel is dat men het perspectief van de potentiële overtreder hanteert ter bepaling van het type bestraffing, zodat een straf ook daadwerkelijk als zodanig wordt beleefd (zie HH&G). De bestuurlijke boete is momenteel de meest toegepaste sanctie tegen bedrijven en op tal van beleidsterreinen heeft de bestuurlijke boete de strafrechtelijke sanctionering verdrongen. In dit onderzoek willen we nagaan in hoeverre bedrijven die een bestuurlijke boete opgelegd krijgen, beter of juist minder goed de regels na gaan leven dan wanneer wordt gekozen voor een ander aanpak (bijvoorbeeld een waarschuwing).

Uit een bestand met onder toezichtgestelden (bedrijven) wordt een steekproef getrokken en deze wordt at random verdeeld over twee groepen. Alle geselecteerde bedrijven worden onaangekondigd bezocht door een inspecteur en een bepaalde regel (het liefst een die met enige regelmaat wordt overtreden) wordt gecontroleerd. Overtreders uit groep 1 krijgen een bestuurlijke boete terwijl overtreders uit groep 2 een waarschuwing krijgen. Vervolgens worden dezelfde bedrijven na een bepaalde periode (bijvoorbeeld een

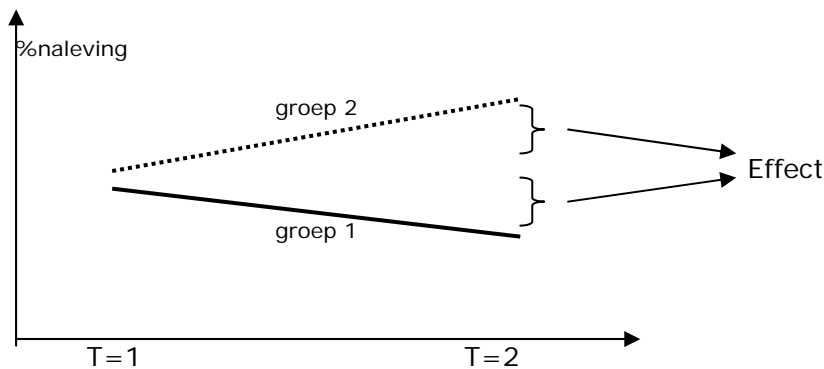
half jaar) nog eens bezocht en wordt dezelfde regel wederom gecontroleerd. Ook deze opzet kunnen we schematisch weergeven:



T



Een mogelijke uitkomst van dit experiment kan vervolgens grafisch gepresenteerd worden:



Dit experiment kan ook uitgevoerd worden door in groep 2 alleen de naleving te controleren en verder niets te ondernemen, mocht er sprake is van een overtreding. Daarnaast kan dit experiment worden herhaald voor diverse regels of met andere doelgroepen.

Wat het oplevert:

De uitkomsten van studie 2 kunnen worden gebruikt voor de ontwikkeling van meer effectieve toezicht-, en handhavingstrategieën. Door systematisch te variëren in de wijze waarop wordt gereageerd op geconstateerde overtredingen en vervolgens te controleren in hoeverre normadressaten zich in de loop der tijd aan deze regel houden, ontstaat inzicht in het relatieve belang van deze aanpakken. Wellicht levert een waarschuwing in bepaalde omstandigheden uiteindelijk meer op dan het uitschrijven van bestuurlijke boetes. Het is daarnaast interessant om op het niveau van de normadressaat te onderzoeken hoe een waarschuwing of een boete wordt beleefd. Dit doen we in het onderzoeksprogramma HH&G.

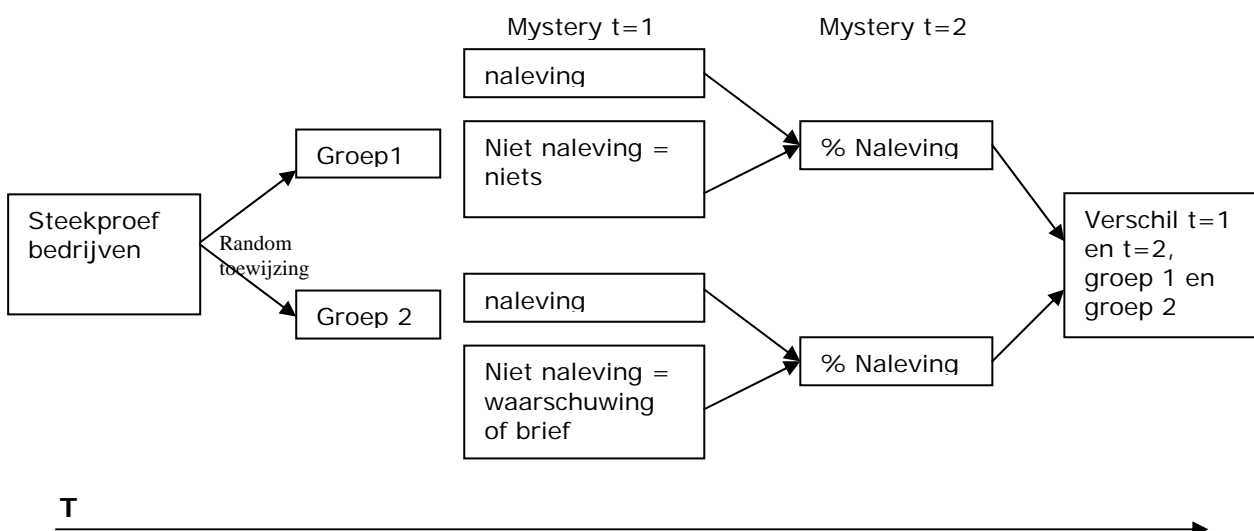
Studie 3

Titel: De effecten van zelfregulering op naleving
 Raamwerk: F
 Type: Veldexperiment – Mystery shopping

Burgers, bedrijven en organisaties riepen de afgelopen jaren dat ze te veel last hebben van bemoeienissen van toezichthouders, door de 'druk(te)' die ze veroorzaken, de administratieve lasten die daarmee gepaard gaan en omdat lang niet altijd duidelijk is hoe toezicht en handhaving bijdraagt aan het gemeenschappelijk belang. Toezicht moest kortom anders, efficiënter en effectiever en hieraan gerelateerd ging het vorige Kabinet er vanuit dat inspecties zoveel mogelijk gebruik gingen maken van controlesystemen die in de onder toezicht staande organisaties en ketens zelf al bestaan. Er werd met andere woorden gepleit voor meer trust-and-hope regulation, niet alleen in Nederland maar ook in andere westerse landen (Shover, 2008). Zelfregulering valt hier ook onder en is goeddeels gebaseerd op vertrouwen in onder toezichtgestelden. Zelfreguleringsystemen bestaan uit een stelsel van zelfreguleringsactiviteiten. In aansluiting op de reguleringsketen zoals die in de literatuur is beschreven (Dorbeck-Jung, Oude Vrielink & Reussing, 2006) bevat dit stelsel samenhangende reguleringsactiviteiten die gericht zijn op de normstelling, de uitvoering (invoering en toepassing) en de handhaving (toezicht en sanctionering) van regels. Een relevante vraag is in hoeverre zelfregulering effectiever is dan andere vormen van regulering, controle en sanctionering. In deze studie willen we dit met behulp van mystery-shoppers bepalen.

Een goed voorbeeld van een mystery-shop onderzoek dat in opdracht van het WODC is uitgevoerd, is de naleving van de leeftijdsclassificaties van Kijkwijzer en PEGI, zodat geen audiovisuele producten (dvd, game of bioscoopkaartje) verkocht/verhuurd of vertoond worden aan jongeren die volgens de leeftijdsclassificatie daar nog te jong voor zijn. Daartoe hebben 11-jarige en 15-jarige mysteryshoppers in totaal 528 aankoop pogingen gedaan om audiovisuele producten te kopen met de leeftijdsclassificaties 12 jaar en 16 jaar (waarbij de 11-jarigen producten probeerden te kopen of te huren met de leeftijdsclassificatie 12 jaar en 16 jaar en 15-jarigen producten probeerden te kopen of te huren met de leeftijdsclassificatie 16 jaar). Er zijn verschillende aanbieders van audiovisuele media bezocht, waaronder warenhuizen, speelgoedwinkels, cd/dvd winkels, gameshops, videotheken bibliotheken en bioscopen. Uit het onderzoek bleek dat jongeren doorgaans weinig moeilijkheden ondervonden bij het aanschaffen van een dvd, game of bioscoopkaartje voor een film waarvoor ze te jong zijn. De correcte naleving was in totaal 14%.

Het is nu interessant om deze exercitie nog eens te herhalen, maar dan met een wat meer experimenteel karakter. We nemen hier de naleving van de leeftijdsclassificatie als voorbeeld, maar het mag ook een andere regel zijn waarvan wordt verondersteld dat zelfregulering een goede manier is om te zorgen voor naleving.



De figuur laat zien dat het nalevingspercentage op twee momenten wordt gemeten voor beide groepen. Voor groep 1 geldt dat zij op geen enkele manier merken dat er sprake is van controle terwijl groep 2 wel wordt geconfronteerd met geconstateerde overtredingen.

Wat het oplevert:

De uitkomsten van studie 3 maken inzichtelijk in hoeverre zelfregulering in een bepaalde sector zorgt voor acceptabele nalevingsniveau's. Door in groep 2 wel te handhaven, is het mogelijk om de relatieve effectiviteit van zelfregulering te bepalen t.o.v. de alternatieve aanpak. Uiteraard gelden die resultaten alleen binnen de sector waar het onderzoek zich afspeelt.

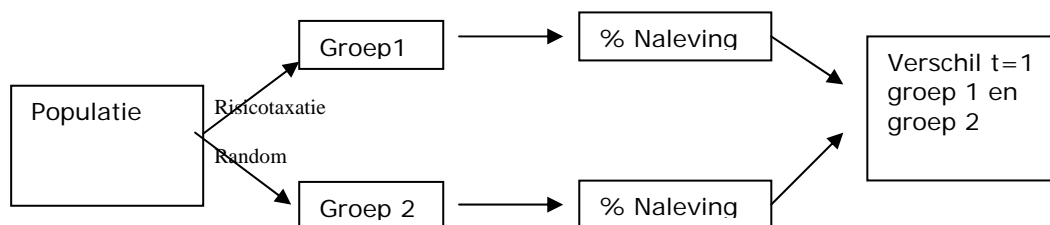
Studie 4

Titel: De effectiviteit van risicogestuurd toezicht (gebruik random controles)

Raamwerk: F / G

Type: Veldexperiment

Op diverse beleidsterreinen wordt gewerkt aan toezicht en handhaving op basis van risicotaxatie. Er wordt dan steeds minder opgetreden op basis van geconstateerde overtredingen, maar op basis van de inschatting van risico's. De hiertoe ontwikkelde modellen zijn in belangrijke mate geïnspireerd door methoden van risk assessment in het bedrijfsleven (Huisman, 2007). Een belangrijke vraag is hoe betrouwbaar en valide deze risicotaxaties zijn en of ze in staat zijn de grootste gevaren te neutraliseren. In deze studie willen we de resultaten van toezichtsactiviteiten op basis van een risicotaxatie afzetten tegen toezicht op basis van willekeur (random controles).



Wat het oplevert:

De uitkomsten van studie 4 maken inzichtelijk in hoeverre inspecties o.b.v. een bepaalde risicotaxatie meer regelovertreding detecteren dan volstrekt willekeurige inspecties. Door het organiseren van (een beperkt aantal) random inspecties (zie ook studie 7), ontstaat er een benchmark waarmee meer geavanceerde inspectiemethoden vergeleken kunnen worden. De opbrengst is het relatieve effect van de methode t.o.v. willekeur.

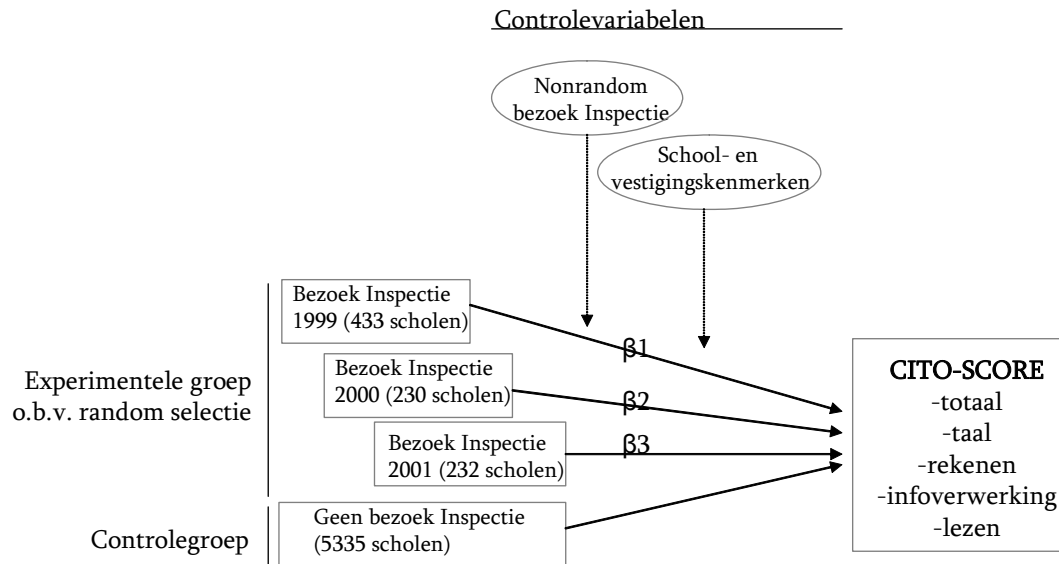
Studie 5

Titel: Het vaststellen van effecten op basis van een registratie

Raamwerk: F / G

Type: Secundaire analyse

Een voorbeeld van een natuurlijk experiment op basis van registraties geven Luginbuhl, Webbink & De Wolf (2007). In dit onderzoek is vastgesteld wat het effect is van schoolbezoeken door de Inspectie van het Onderwijs op de prestaties van leerlingen. De resultaten laten zien dat de scores op de CITO-toets verbeteren in het eerste en tweede jaar na het bezoek van de Inspectie. Dit geldt zowel voor de scores op de totale CITO-toets als voor de scores op de deoltoetsen voor taal, informatieverwerking en begrijpend lezen. Voor rekenen houdt het inspectie-effect zelfs tot vier jaar na het inspectiebezoek aan. In de analyses wordt gecontroleerd voor allerlei andere aspecten die mogelijk van invloed zijn op de prestaties van kinderen, zoals eerder uitgevoerde inspectiebezoeken en school- en vestigingskenmerken. In de onderstaande figuur is het onderzoeksdesign grafisch weergegeven.



In deze studie is gebruik gemaakt van een natuurlijk experiment. Het vergelijken van wel en niet bezochte scholen is minder zinvol, omdat de selectie ten behoeve van reguliere inspecties niet at random tot stand komt. Scholen die in het verleden lager op kwaliteitscriteria scoorden, hebben immers een grotere kans om bezocht te worden dan andere scholen. Ten behoeve van het Onderwijsverslag dat de Inspectie jaarlijks opstelt onderzoekt deze Inspectie honderden scholen, die wél op basis van toeval in de selectie terecht komen. Het zijn deze random steekproeven die door de onderzoekers zijn gebruikt als experimentele groep, terwijl alle andere scholen de controlegroep vormen. Het verschil met een gecontroleerd experiment is dat de toewijzing in dat geval niet door de onderzoeker wordt gedaan, maar het gevolg is van toevallige elementen in de regelgeving, in dit geval de verplichting om jaarlijks een verslag te publiceren.

In studie 5 willen we een soortgelijke exercitie uitvoeren maar dan op een ander domein.

Nb: We hebben hier vijf mogelijke experimentele studies gepresenteerd. Graag bespreken we met de participanten van EvT de verdere inrichting van deze studies omdat het moeilijk voor ons is te bepalen wat de mogelijkheden zijn.

4.2 Niet experimentele onderzoeken

Studie 6

Titel: Zoektocht naar valide en betrouwbare bronnen maatschappelijk effecten
Raamwerk: -
Type: Inventarisatie

Deze inventarisatie dient in kaart te brengen welke bronnen kunnen worden aangewend om uitspraken te doen over de maatschappelijke effecten van toezicht en handhaving. In het rapport van Winter en De Ridder worden voorbeelden genoemd op het terrein van verkeersveiligheid (ongevallenregistraties) en het onderwijs (Citoscores e.d.). Sommige registraties zullen de inspecties zelf in huis hebben, terwijl andere worden beheerd door organisaties buiten de toezichtwereld. Naast een inventarisatie van bronnen binnen diverse kleursporen dient ook bepaald te worden of de gegevens valide en betrouwbaar zijn.

Wat het oplevert:

Met de uitkomsten van dit onderzoek kunnen inspecties makkelijker en met meer zekerheid gebruik maken van interne en externe gegevensbestanden over maatschappelijke effecten. Hiermee proberen we te voorkomen dat gegevens 'ongebruikt' op de plank blijven liggen maar juist worden ingezet t.b.v. het vaststellen van effecten.

Studie 7

Titel: De mogelijkheden en kosten van structurele random audits
Raamwerk: -
Type: Kosten Baten Analyse

In studie 4 propageerden we het gebruik van random audits omdat het een goede manier is om het relatieve effect van risicogestuurde toezichtstrategieën te kwantificeren. Op basis van dit type onderzoek kunnen bestaande strategieën worden gecontroleerd en verbeterd. Een vaak gehoord argument om geen gebruik te maken hiervan is het doelmatigheidsargument. Het is niet efficiënt om mensen en middelen willekeurig in te zetten t.b.v. controles en inspecties. In deze studie willen we onderzoeken hoeveel random controles uitgevoerd moeten worden om verschillen t.o.v. reguliere inspectiemethoden aan te tonen. Het aantal uit te voeren random controles is onder meer afhankelijk van de prevalentie van de overtreding van een bepaalde regel en daarom zal de uitkomst variëren. Voor de ene regel zal het doelmatigheidsargument gelden terwijl dat voor andere regels juist niet geldt en het verstandig is om random audits in te bouwen in de toezichtspraktijk. Naast het bepalen van random inspectieomvang, dient in deze studie te worden bepaald op welke wijze deze georganiseerd kunnen worden.

Studie 8

Titel: Reconstructie van wetgeving en toezicht X
Raamwerk: E
Type: Reconstructie van de
Discipline: Bestuurskunde, bedrijfskunde, economie

Winter en De Ridder gaven aan dat het belangrijk is om te expliciteren hoe het gedrag van de normadressaat wordt beïnvloed door de ingezette interventie en hoe dit uiteindelijk resulteert in het beoogde maatschappelijke effect. Bij het ontwikkelen van een onderzoeksdesign is het goed te weten hoe de beleids- of interventietheorie in elkaar steekt. Soms bestaat deze al maar meestal moet deze nog ontwikkeld worden. Door oorzaak-gevolg relaties uit te werken wordt vooraf duidelijk hoe effecten tot stand

zouden moeten komen. Door een goede reconstructie komen ook positieve en negatieve neveneffecten boven tafel. Om te laten zien hoe een reconstructie uitgevoerd dient te worden, willen we er een programmeren.