

Handleiding Doelmatigheids- en doeltreffendheids- onderzoek

2023



Inhoud

1. Algemene inleiding | 4

- 1.1 Doel, gebruik en opzet van de handleiding | 4
- 1.2 Kernbegrippen | 5
- 1.3 Onderzoeksontwerp: de kern van je onderzoek | 6

2. Kenmerken rekenkameronderzoek | 7

- 2.1 Inleiding | 7
- 2.2 Wat onderzoeken we? | 7
- 2.3 Tot welke resultaten leidt ons onderzoek? | 9

3. Doelstelling, probleemstelling en vraagstelling | 10

- 3.1 Inleiding | 10
- 3.2 Doelstelling | 10
- 3.3 Probleemstelling | 11
- 3.4 Onderzoeksvragen | 12

4. Beschrijvende onderzoeksvragen | 13

- 4.1 Inleiding | 13
- 4.2 Varianten voor beschrijvende vragen | 13
- 4.3 Methoden: hoe zijn beschrijvende vragen in eerder onderzoek beantwoord? | 15
- 4.4 Conclusies | 17

5. Toetsende onderzoeksvragen | 18

- 5.1 Inleiding | 18
- 5.2 Wat is een toetsende vraag? | 18
- 5.3 Toetsen aan een norm | 19
- 5.4 Methoden: hoe zijn toetsende vragen in eerder onderzoek beantwoord? | 23
- 5.5 Conclusies en oordelen | 24

6. Verklarende onderzoeksvragen | 25

- 6.1 Inleiding | 25
- 6.2 Wat is een verklarende vraag? | 25
- 6.3 Methoden: hoe zijn verklarende vragen in eerder onderzoek beantwoord? | 31
- 6.4 Soorten conclusies | 31

7. Onderzoeksubjecten en -varianten | 32

- 7.1 Inleiding | 32

- 7.2 Middelenv33
- 7.3 Activiteiten | 33
- 7.4 Prestaties | 34
- 7.5 Effecten | 35
- 7.6 Onderzoek naar prestaties en effecten van derden bij gedecentraliseerd beleid | 37
- 7.7 Indicatoren voor resultaten van beleid | 38
- 7.8 Onderzoekvarianten | 38

8. Algemene normen voor DDO | 44

- 8.1 Inleiding | 44
- 8.2 Basisnormen beleidsvoorbereiding | 46
- 8.3 Basisnormen beleidsuitvoering | 50
- 8.5 Basisnormen gebruik beleidsinformatie door gecontroleerden | 55
- 8.6 Normen voor de kwaliteit van derdeninformatie voor gebruik in eigen onderzoekv58
- 8.7 Normen voor de beoordeling van de overheidsrol in beleidsnetwerken | 66
- 8.8 Normen voor toezicht op naleving en handhaving van EU wet- en regelgeving | 69
- Bijlage: uitwerking beoordelen van beleidsdoelen | 71

9. Methoden en technieken van dataverzameling en -analyse | 74

- 9.1 Inleiding | 74
- 9.2 Methoden en technieken van dataverzamelingv74
- 9.3 Methoden en technieken van data-analysev84

10. Bevindingen, conclusies en bestuurlijke uitspraken | 103

- 10.1 Inleiding | 103
- 10.2 Bevindingen | 103
- 10.3 Feitelijke conclusies | 105
- 10.4 Bevindingen en feitelijke conclusies: de nota van bevindingen | 105
- 10.5 Bestuurlijke uitspraken | 107

11. Methodische bijlage | 109

- 11.1 Inleiding | 109
- 11.2 Tips en voorbeelden bij het schrijven van een methodische bijlage | 110
- 11.3 Definities | 117

12. Literatuur en inspiratie | 118

- 12.1 International Standards of Supreme Audit Institutions | 118
- 12.2 Algemeen | 118
- 12.3 Beleidsvoorbereiding | 118

13. Verklarende begrippenlijst | 119

1. Algemene inleiding

1.1 Doel, gebruik en opzet van de handleiding

Deze handleiding geeft inhoudelijke ondersteuning bij het doen van doelmatigheids- en doeltreffendheidsonderzoek door rekenkamers. In deze handleiding gebruiken we veel voorbeelden van eerder onderzoek van de Algemene Rekenkamer. Daarmee illustreren we hoe een aanpak er concreet uit kan zien. Besef wel dat ieder onderzoek zijn eigen context en dynamiek kent.

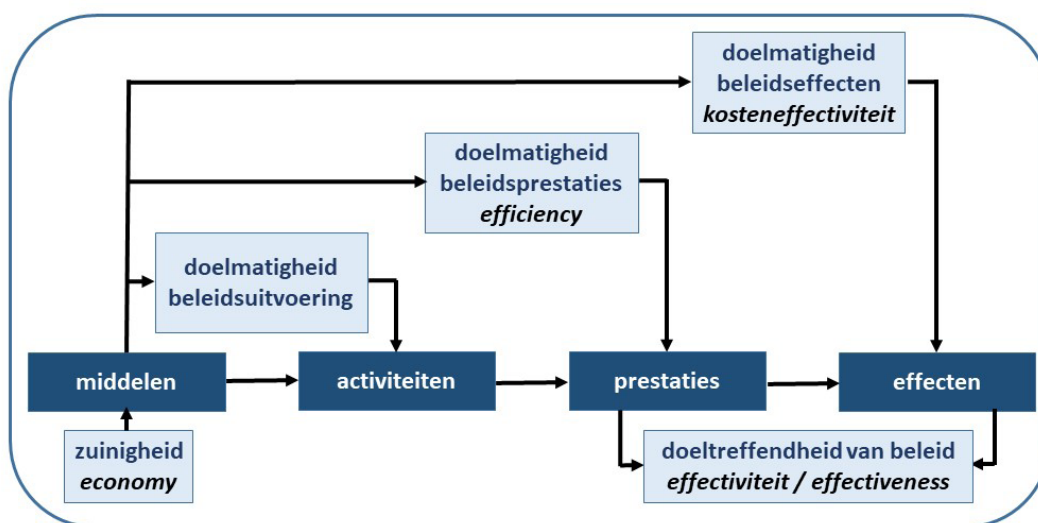
Met deze handleiding willen we onderzoekers en andere betrokkenen ondersteunen bij rekenkameronderzoek en de beoordeling daarvan. Onderzoek is creatief maatwerk, dus je kunt het niet in regels vatten. Tegelijk gelden bij het voorbereiden, uitvoeren en rapporteren van onderzoek wel een aantal basale principes (zie ook [ISSAI 300](#), Performance Audit Principles). Het is aan de onderzoekers om die principes op een verstandige wijze toe te passen, niet als een keurslijf, maar als een welkome hulp om de kwaliteit van je onderzoek te waarborgen. Daarnaast kun je je voordeel doen met de vele voorbeelden en toelichtingen die in deze handleiding zijn opgenomen.

In de hoofdstukken 3 t/m 9 vind je ondersteuning bij het vormgeven (design) en het uitvoeren van je onderzoek. Hoofdstuk 10 (formuleren van bevindingen, conclusies en aanbevelingen) en hoofdstuk 11 (methodische bijlage) kun je gebruiken in de rapportage- en publicatiefase van je onderzoek.

1.2 Kernbegrippen

Doelmatigheids- en doeltreffendheidsonderzoek, internationaal aangemerkt als performance audit, heeft betrekking op onderzoek gerelateerd aan *economy*, *efficiency* en *effectiveness*. Kernbegrippen zijn dus zuinigheid (*economy*), doelmatigheid (*efficiency*) en doeltreffendheid (*effectiveness*).

Figuur 1.1 Kernbegrippen in hun samenhang



In onderzoek naar **zuinigheid** vragen we ons af in hoeverre men de kosten van de verworven middelen minimaliseert. Randvoorwaarden daarbij zijn dat de middelen tijdig beschikbaar, van gepaste kwaliteit en in de gepaste hoeveelheid voorhanden zijn, en verworven tegen de best mogelijke prijs.

In onderzoek naar **doelmatigheid** vragen we ons af in hoeverre de beschikbare middelen (personeel, materieel, geld) optimaal worden benut. We onderscheiden daarbij drie soorten doelmatigheid: de relatie tussen de middelen en de verrichte activiteiten om het beleid uit te voeren (doelmatigheid van de beleidsuitvoering), de relatie tussen de middelen en de geleverde prestaties (doelmatigheid van de beleidsprestaties) en de relatie tussen de middelen en de gerealiseerde effecten (doelmatigheid van de beleidseffecten). Met andere woorden: de mate waarin de uitvoering, prestaties en/of effecten met de inzet van minder middelen gerealiseerd hadden kunnen worden, of de mate waarin een betere uitvoering, méér prestaties en/of méér effecten verwezenlijkt hadden kunnen worden met dezelfde inzet van middelen.

In onderzoek naar **doeltreffendheid** vragen we ons af in hoeverre *met het beleid* de gestelde doelen, respectievelijk de verwachte resultaten (prestaties en/of effecten) zijn behaald.

In deze handleiding gaan we in op doelmatigheid en doeltreffendheid. Het principe van zuinigheid kan worden betrokken in een doelmatigheidsonderzoek. In hoofdstuk 7 gaan we dieper in op de verschillende onderzoeksobjecten en -varianten.

1.3 Onderzoeksontwerp: de kern van je onderzoek

Het formuleren van een onderzoeksplan begint in de regel bij de doelstelling en probleemstelling en de onderzoeksvragen: wat willen we onderzoeken en waarom? Daarnaast denk je na over normen die je gaat hanteren (in een toetsend onderzoek), over de methoden van dataverzameling en hoe je de data gaat analyseren. Uiteindelijk moet dit onderzoeksontwerp een samenhangend geheel zijn. Het type onderzoeksvraag is heel bepalend voor de overige fasen van het onderzoek. Daarom is het belangrijk om goed na te denken over de formulering van je onderzoeksvragen.

Uit een analyse van gepubliceerde rekenkamerrapporten blijkt dat in ons onderzoek in feite drie soorten vragen aan de orde komen: beschrijvende vragen (hoofdstuk 4), toetsende vragen (hoofdstuk 5) en verklarende vragen (hoofdstuk 6). In de meeste onderzoeken wordt gebruik gemaakt van een combinatie van beschrijvende, toetsende en verklarende vragen.

Je onderzoeksvraag bepaalt welke aanpak of methoden het best kunnen worden gebruikt. De combinatie van onderzoeksvraag en aanpak is doorslaggevend voor het soort conclusies dat kan worden getrokken. Als in het onderzoek alleen beschrijvende onderzoeksvragen worden opgenomen, kun je geen oordelen geven over wat je tegenkomt. Als je wel tot een oordeel wilt komen, moet je dus een toetsende onderzoeksvraag in je onderzoek opnemen. In het algemeen geldt dat de onderzoeksvraag, de gehanteerde methode en het type conclusies dat je op basis van je onderzoek kunt trekken, een 'logische' combinatie vormen.

2. Kenmerken rekenkameronderzoek

2.1 Inleiding

Rekenkameronderzoek volgt in principe de normale stappen van een wetenschappelijk onderzoek. Het belangrijkste verschil is dat er in een rekenkameronderzoek in de regel wordt gewerkt met normatieve onderzoeksvragen, terwijl deze in wetenschappelijk onderzoek slechts bij uitzondering worden gebruikt. Dit komt omdat de Algemene Rekenkamer en andere rekenkamers controle- of auditinstituten zijn en het onderzoek daarom in principe is gericht op het geven van oordelen. En een oordeel kan alleen worden gegeven als er aan een norm is getoetst.

In dit hoofdstuk beschrijven we de typische kenmerken van rekenkameronderzoek.

2.2 Wat onderzoeken we?

We onderzoeken vrijwel altijd vastgesteld beleid. Heel soms onderzoeken we processen of producten die nog niet zijn afgerond, en dan meestal op verzoek van de betrokken bewindspersoon, maar we vellen geen oordeel over voorgenomen beleid. Dat is voorbehouden aan de Tweede en Eerste Kamer. De Raad van State toetst de kwaliteit en uitvoerbaarheid van wetten en voorstellen en bekijkt of deze in overeenstemming zijn met andere wettelijke bepalingen en (internationale) verdragen. Rekenkamers kunnen wel achteraf de voorbereiding van het beleid onderzoeken.

2.2.1 Typen vraagstellingen

In de meeste gevallen is het onderzoek van de Algemene Rekenkamer erop gericht niet alleen feitelijke conclusies te trekken (“dit is hoe het is”), maar ook om een bestuurlijk oordeel te geven (“dit is wat we ervan vinden”) en vervolgens

aanbevelingen te doen ("dit is hoe het beter kan").

Typische Rekenkamer-onderzoeksvragen zijn:

- Hoeveel heeft het beleid gekost (of: heeft het beleid gekost wat de minister dacht dat het zou kosten)?
- Wat zijn de resultaten van het beleid (prestaties, effecten) (of: heeft het beleid opgeleverd wat de minister dacht dat het zou opleveren)?
- Had het goedkoper gekund?
- Is de Tweede Kamer hierover geïnformeerd met objectieve en betrouwbare informatie?
- Wat zijn de oorzaken van het tekortschieten van doeltreffendheid/ doelmatigheid/ goede informatievoorziening aan de Tweede Kamer?

In hoofdstuk 4, 5 en 6 gaan we uitgebreid in op de verschillende typen onderzoeksvragen. In hoofdstuk 7 gaan we dieper in op de verschillende onderzoeksobjecten en -varianten.

2.2.2 Focusonderzoek

Soms doen we onderzoek om alleen de feiten boven tafel te krijgen binnen een actuele discussie. We rapporteren dan ook alleen over die feiten (in hun context) en formuleren geen conclusies en/of aanbevelingen. Dit type onderzoek wordt focusonderzoek genoemd. In een focusonderzoek wordt een beschrijvende onderzoeksvraag gesteld. Hieronder volgen een aantal voorbeelden van centrale onderzoeksvragen:

- Blijven aangiften met opsporingsindicatie liggen en hoe lang duurt dat (onderzoek [Aangiften bij de politie](#))?
- De minister van EZ claimde al in 2016 dat de kosten van windenergie op zee met 40% zijn afgenomen. In 2018 meldt de minister van EZK bovendien dat in 2022 het eerste windpark zonder subsidie tot stand zal worden gebracht. In dit onderzoek zijn we nagegaan of deze beweringen kloppen (onderzoek [Kosten windenergie op zee](#)).
- Hoe heeft het percentage afwijzingen voor Wlz-indicaties zich in de periode 2015–2017 ontwikkeld en zien we hierbij verschillen tussen gemeenten en tussen groepen aanvragers? Hoe kunnen we deze ontwikkeling en eventuele verschillen duiden (onderzoek [Toegang tot langdurige zorg](#))?
- Wat zijn de gevolgen van de gekozen governancestructuur voor de totstandkoming en voortgang van de gekozen wederopbouwprojecten op Sint-Maarten, te financieren uit het trustfonds (onderzoek [Nederlandse bijdrage aan de wederopbouw van Sint-Maarten](#))?

2.3 Tot welke resultaten leidt ons onderzoek?

Rekenkameronderzoek levert verschillende uitkomsten op. In veel onderzoeken gaan we na of de praktijk (de werkelijke situatie) voldoet aan een of meer normen (de gewenste situatie). Voor het beschrijven van de uitkomsten van dergelijk onderzoek gebruiken we verschillende begrippen.

Begrip	Voorbeeld
Onderzoeksvraag	Voldoen de gerealiseerde wachttijden in de jeugdhulp aan de daaraan door de minister gestelde normen?
Normen	De maximale wachttijd die in wet- en regelgeving is vastgelegd mag niet overschreden worden. In het geval van opvoedhulp ligt die maximale wachttijd op 6 weken.
Feiten	35% van de aanvragen voor opvoedhulp wordt na 6 weken of later in behandeling genomen.
Bevinding: antwoord op de vraag of aan de norm is voldaan	In 35% van de door ons bestudeerde gevallen is de maximale wachttijd overschreden voor de opvoedhulp.
(Feitelijke) Conclusie: antwoord op de onderzoeksvraag	De minister voldoet niet aan alle maximale wachttijden in de jeugdhulp (waaronder die voor de opvoedhulp).
Bestuurlijke uitspraak/ oordeel: waardeoordeel over een bepaalde praktijk of situatie; geeft aan hoe erg we het vinden	De wachttijden voldoen <i>onvoldoende</i> aan de door de minister daaraan gestelde normen ('onvoldoende' is hier het waardeoordeel). <i>Let op:</i> een bestuurlijk oordeel is alleen opgenomen in een eindrapport (niet in een nota van bevindingen) en moet de instemming hebben van het college. Ook een aanbeveling is een bestuurlijke uitspraak.

Zie ook hoofdstuk 10.

3.

Doelstelling, probleemstelling en vraagstelling

3.1 Inleiding

De doelstelling, probleemstelling en vraagstelling zijn bepalend voor de opzet, de scope, de uitvoering en de uitkomsten van een onderzoek. In een onderzoek kan de neiging ontstaan om 'alles' te onderzoeken. Het formuleren van een goed afgebakende doelstelling en probleemstelling disciplineert en helpt te voorkomen dat een onderzoek hierdoor vertraagt of zelfs ontspoot. De ervaring leert dat deze discipline niet vanzelfsprekend is. Deze fase is dus een belangrijk onderdeel van het onderzoek dat de nodige aandacht verdient. In hoofdstuk 7 vind je een beschrijving van de verschillende onderzoeksobjecten en -varianten die je kunnen helpen bij de afbakening van je onderzoek.

Daarnaast is het belangrijk om in deze fase van het onderzoek zorgvuldig na te gaan in hoeverre de minister of het bestuursorgaan verantwoordelijk kan worden gehouden voor het probleem dat je wilt onderzoeken en in hoeverre de minister of het bestuursorgaan voldoende bevoegd en in staat is om het probleem op te lossen.

3.2 Doelstelling

De doelstelling van een onderzoek geeft aan wat de onderzoeker met een onderzoek wil bereiken. Het gaat hier dus om de vraag waarom het onderzoek gedaan wordt. Voor rekenkameronderzoek geldt in het algemeen dat met de doelstelling wordt aangegeven welk (maatschappelijk) effect men wil bereiken met het onderzoek. Vaak gaat het daarbij in eerste instantie om het bewerkstelligen van een bepaalde actie van de verantwoordelijke bewindspersoon en/of de betrokken instelling. We spreken in dat geval van een 'handelingsgerichte' doelstelling.

Voorbeelden van gangbare doelstellingen voor rekenkameronderzoek zijn 'het bevorderen van de kwaliteit van het uitvoeringsproces bij een RWT' of 'het leveren van een bijdrage aan de verbetering van de doelmatigheid van een bepaalde dienst'. In het rapport [Stimulering van duurzame energieproductie](#) wordt de doelstelling van het onderzoek expliciet omschreven: "bijdragen aan een (nog) beter functionerende sde+, waardoor de kans toeneemt dat de duurzame energiedoelen voor 2020 en 2023 op een kosteneffectieve manier worden gehaald".

Een ander voorbeeld is het rapport [Basisregistraties](#) waarin de doelstelling luidt: "bijdragen aan de oplossing van (mogelijke) knelpunten die zich in het stelsel of bij de afzonderlijke basisregistraties voordoen vanuit het perspectief van de burger en vanuit de invalshoek van fraudebestrijding."

3.3 Probleemstelling

De probleemstelling is de centrale vraag waarop door middel van het onderzoek een antwoord gezocht wordt. De probleemstelling geeft op hoofdlijnen aan wat er onderzocht gaat worden. Het te onderzoeken probleem wordt in vraagvorm geformuleerd, bijvoorbeeld: faciliteert het beleid de lokale beleidsuitvoerders voldoende om de gewenste resultaten te boeken, is het beleid toereikend om gesignaleerde problemen op te lossen en welke knelpunten doen zich daarbij voor? De probleemstelling is dus 'kennisgericht'.

De probleemstelling heeft de functie de complexiteit van een onderzoeksterrein beheersbaar te maken door een afbakening te maken die hanteerbaar is, maar die tegelijkertijd ook weer niet te veel simplificeert. De probleemstelling stuurt daarmee de onderzoeker tijdens de uitvoering van het onderzoek. Let er ook op dat de probleemstelling zo nodig geografisch en in tijd is afgebakend.

De probleemstelling kan niet los worden gezien van de doelstelling van het onderzoek. Het antwoord op de vraag of vragen waaruit de probleemstelling is opgebouwd moet, samen met de daaruit voortvloeiende aanbevelingen, een bijdrage leveren aan de realisatie van de doelstelling.

De koppeling met de doelstelling is meer dan een verplicht nummer en verdient serieuze aandacht. Als een onderzoek de pretentie heeft een bijdrage te leveren aan de doelmatigheid van een bepaald beleid, moeten de probleemstelling en de onderzoeksopzet daarop toegesneden zijn.

Uiteraard dient er ook een inhoudelijk verband te bestaan tussen doelstelling en probleemstelling: beide moeten betrekking hebben op hetzelfde onderwerp. De ervaring leert dat zelfs tegen deze vanzelfsprekende voorwaarde soms wordt gezondigd.

In het rapport [Stimulering van duurzame energieproductie](#) sluit de probleemstelling ('de centrale vraag') goed aan op de doelstelling: "Moet er iets worden veranderd aan de sde+ om de subsidieregeling effectiever en efficiënter te maken en zo ja, wat zou er dan precies moeten worden aangepast?"

In het rapport [Basisregistraties](#) luidt de probleemstelling: "Welke knelpunten ervaren betrokkenen in (het stelsel van) de basisregistraties vanuit het perspectief van de burger en ten aanzien van de bijdrage aan fraudebestrijding, worden deze knelpunten aangepakt en hoe worden in dit verband mogelijke risico's beheerst?" Ook deze probleemstelling sluit goed aan bij de doelstelling van het onderzoek.

3.4 Onderzoeksvragen

Onderzoeksvragen zijn een uitwerking van de probleemstelling en hebben daarmee dus een logische relatie. In samenhang horen de onderzoeksvragen een antwoord te geven op de probleemstelling.

Een goede uitwerking in onderzoeksvragen voorkomt dat het onderzoek ruimer wordt dan de probleemstelling. Het is verleidelijk om bij de onderzoeksvragen nog wat extra interessante vragen op te nemen, omdat die mogelijk 'wat belangrijks' opleveren. Als echter de probleemstelling zowel voldoende breed als voldoende afgebakend is, is deze verholde uitbreiding van het onderzoek niet nodig. Omdat onderzoek in veel gevallen een niet lineair karakter heeft en er sprake kan zijn van voortschrijdend inzicht, kan het nodig zijn in de loop van het proces een onderzoeksvraag aan te passen, bijvoorbeeld omdat je anders de probleemstelling niet kunt beantwoorden.

In rekenkameronderzoek kunnen verschillende soorten onderzoeksvragen worden onderscheiden (zie hoofdstuk 1).

Enkele tips voor het opstellen van onderzoeksvragen:

- Breng structuur aan in de onderzoeksvragen, bijvoorbeeld door in elke vraag slechts één onderdeel van het probleem aan te pakken. Elk onderdeel kan dan apart worden onderzocht en in samenhang met de andere onderdelen bijdragen aan de oplossing van de probleemstelling.
- Controleer of alle onderzoeksvragen samen tot totale beantwoording van de probleemstelling kunnen leiden.
- Stel niet te veel onderzoeksvragen. Een probleemstelling is meestal met twee tot zes vragen af te dekken. Als er meer rijzen stel jezelf dan de vraag of de onderzoeksvragen niet verder gaan dan de probleemstelling beoogt, dan wel of de probleemstelling niet te breed is en een te uitgebreid onderzoek vergt.
- Formuleer onderzoeksvragen niet te gedetailleerd. Let op of twee vragen in werkelijkheid wellicht als één vraag te beschouwen zijn.

4.

Beschrijvende onderzoeksvragen

4.1 Inleiding

Nadat je de doel- en probleemstelling hebt geformuleerd, vertaal je dit naar een vraagstelling voor je onderzoek. De vraagstelling hangt af van het soort kennis waar je naar op zoek bent om je doelstelling te bereiken. Daarom is het belangrijk om goed na te denken over de formulering van je onderzoeksvragen. Uit een analyse van gepubliceerde rekenkamerrapporten blijkt dat in ons onderzoek in feite drie soorten vragen aan de orde komen: beschrijvende vragen, toetsende vragen en verklarende vragen. In de meeste onderzoeken wordt gebruik gemaakt van een combinatie van beschrijvende, toetsende en verklarende vragen. Beschrijvende vragen richten zich over het algemeen op basiskennis die je nodig hebt alvorens je eventueel complexere vragen (zoals een toetsende of verklarende vraag) kan beantwoorden. We gaan nu in op het type 'beschrijvende onderzoeksvraag'.

4.2 Varianten voor beschrijvende vragen

4.2.1 Wat is een beschrijvende vraag?

In veel Rekenkameronderzoeken proberen we zo goed mogelijk de werkelijkheid te beschrijven. Dit noemen we beschrijvend onderzoek en daarvoor gebruiken we beschrijvende onderzoeksvragen. Kenmerkend voor beschrijvende vragen is dat ze leiden tot inzicht, maar niet tot een oordeel. We onderscheiden drie soorten beschrijvende onderzoeksvragen: kwalitatief beschrijvende vragen, kwantitatief beschrijvende vragen en feitelijk beschrijvende vragen. In het laatste geval is sprake van feitenonderzoek of focusonderzoek. Bij een kwalitatief beschrijvende vraag is de onderzoeksvraag heel open geformuleerd: hoe zit iets in elkaar? Bij een kwantitatief

beschrijvende vraag zijn we bijvoorbeeld op zoek naar aantallen: hoe vaak komt iets voor of hoeveel geld is ergens mee gemoed? Met feitelijk beschrijvende vragen willen we bekijken of een bepaalde bewering feitelijk klopt: is het zo dat dit waar is? We gaan nu in op deze drie soorten beschrijvende vragen. In paragraaf 4.3 geven we ook voorbeelden van methoden die we hiervoor gebruikt hebben in eerder AR-onderzoek.

In hoofdstuk 7 vind je voorbeelden van beschrijvende vragen in relatie tot doelrealisatie, doelgroepbereik, doeltreffendheid en doelmatigheid.

4.2.2 Een kwalitatief beschrijvende vraag

Kwalitatief beschrijvende vragen zijn relevant als we nog niet veel van een onderwerp weten en nader

willen onderzoeken hoe iets werkt of hoe het is georganiseerd. De vraag heeft daarmee ook duidelijk een begripsvormend karakter. Een paar voorbeelden van kwalitatief beschrijvende vragen zijn:

- Hoe zit het CE-systeem in elkaar (onderzoek [CE-markering](#))?
- Hoe heeft het Nederlandse stelsel voor asiel die tamelijk plotselinge grote groei van asielaanvragen verwerkt (onderzoek [Asielstroom 2014-2016](#))?
- Hoe hebben scholen in het basisonderwijs in 2011 de beschikbare middelen besteed (onderzoek [Passend onderwijs](#))?

Het maken van onderscheid tussen een beschrijvende en een toetsende vraag kan soms lastig zijn, zoals het volgende voorbeeld laat zien. In het onderzoek [Maatschappelijk verantwoord inkopen bij het rijk](#) worden de volgende onderzoeksvragen gesteld:

- Welk deel van de inkoop van het Rijk is daadwerkelijk maatschappelijk verantwoord, afgezet tegen het totale inkoopvolume van het Rijk?
- Hoe rapporteert het Rijk over de realisatie van de maatschappelijk verantwoorde inkoop en wat valt ons hierbij op?

De formulering lijkt in eerste instantie te duiden op een kwalitatief beschrijvend onderzoek. Maar de sleutel ligt hier bij het woord “daadwerkelijk”. In het onderzoek wordt gekeken welke criteria worden gebruikt om te bepalen of iets maatschappelijk verantwoord is. Daarnaast formuleren de onderzoekers een normenkader waar (een rapportage over) maatschappelijk verantwoord inkopen aan zou moeten voldoen. Hieruit blijkt dat het in feite ging om een toetsend onderzoek.

4.2.3 Een kwantitatief beschrijvende vraag

Bij een kwantitatief beschrijvende vraag gaat het erom dat we geïnteresseerd zijn in getallen: hoe vaak komt iets bijvoorbeeld voor of hoeveel geld heeft iets gekost.

Voorbeelden van kwantitatief beschrijvende onderzoeksvragen zijn:

- Hoe hebben de kosten van windenergie op zee zich de afgelopen jaren ontwikkeld (onderzoek [Kosten windenergie op zee](#))?
- Hoe heeft het aantal voorarrestanten zich de afgelopen jaren ontwikkeld (onderzoek [Voorarrest](#))?
- Wat is de formatie, bezetting en inzetbaarheid op alle organisatorische niveaus van de politie (onderzoek [Politie ter plaatse](#))?
- In welke mate worden 22 uitvoeringsorganisaties geraakt in hun taakstelling, in termen van geld (onderzoek [Bezuinigingen op uitvoeringsorganisaties](#))?

De laatste vraag is een specifieke variant van een beschrijvende vraag, waarin het geven van een overzicht voorop staat (bijvoorbeeld uitvoeringsorganisaties, arrondissementen, regio's).

4.2.4 Een feitelijk beschrijvende vraag

Bij een feitelijk beschrijvende vraag wordt een concrete bewering of claim onderzocht aan de hand van feiten: is het zo dat deze bewering klopt? Dit kan ook wel gezien worden als een vorm van “factchecking”. Deze vragen zijn onder meer leidend in ons focusonderzoek.

Voorbeelden van feitelijk beschrijvende vragen zijn:

- Wordt in Nederland vaker dan in andere Europese landen voorlopige hechtenis, een vorm van voorarrest, toegepast (onderzoek [Voorarrest](#))?
- Hoe heeft het percentage afwijzingen voor Wlz-indicaties zich in de periode 2015–2017 ontwikkeld en zien we hierbij verschillen tussen gemeenten en tussen groepen aanvragers (onderzoek [Toegang tot langdurige zorg](#))?

4.3 Methoden: hoe zijn beschrijvende vragen in eerder onderzoek beantwoord?

4.3.1 Inleiding

In hoofdstuk 9 gaan we uitgebreider in op methoden en technieken. Hier noemen we bij wijze van illustratie een aantal voorbeelden van gebruikte methoden in een beschrijvend onderzoek.

Let op: hou er bij de voorbereiding van je onderzoek rekening mee dat relevante data niet altijd beschikbaar zijn of dat de kwaliteit ervan tekortschiet. Zie ook hoofdstuk 8 (Normen voor de kwaliteit van derdeninformatie voor gebruik in eigen onderzoek).

4.3.2 Een kwalitatief beschrijvende vraag

Er zijn verschillende manieren mogelijk om een kwalitatief beschrijvende vraag te

beantwoorden.

- In het onderzoek [Ce-markering](#) is voor de onderzoeksvraag “Hoe zit het CE-systeem in elkaar” gebruikgemaakt van documentanalyse, verdiepende interviews, expertmeetings en werkbijeenkomsten. Met het reconstrueren van een ‘productreis’ hebben we het traject van twee producten gevolgd van producent tot eindgebruiker.
- In het onderzoek [Asielstroom 2014-2016](#) is een data-analyse uitgevoerd op gegevens over het proces dat asielzoekers hebben doorlopen. Aanvullend hierop is informatie opgevraagd bij diverse organisaties die bij het asielproces betrokken zijn. Daarnaast is gebruikgemaakt van parlementaire documentatie en onderzoek van derden.

Ook is het mogelijk om een kwalitatief beschrijvende vraag in kaart te brengen met een *infographic*. Zo is in 2014 voor een parlementaire enquêtecommissie de geschiedenis van de [Hogesnelheidslijn-Zuid](#) in beeld gebracht.

4.3.3 Een kwantitatief beschrijvende vraag

Om kwantitatief beschrijvende vragen te beantwoorden moeten we in de meeste gevallen documenten en/of databestanden opvragen bij de departementen of andere betrokken instanties en die analyseren. Deels kunnen we deze gegevens soms verkrijgen uit openbare bronnen.

- In het onderzoek [Kosten windenergie op zee](#) hebben we voor de vraag hoe de kosten van windenergie zich hebben ontwikkeld, gebruikgemaakt van gegevens van het ministerie van EZK en TenneT en van berekeningen van ECN.
- Voor de ontwikkeling van het aantal voorarrestanten in het onderzoek [Voorarrest](#) is gebruikgemaakt van gegevens uit openbare bronnen, aangevuld met gegevens die we hebben opgevraagd bij de Nationale Politie en DJI.
- Om in beeld te krijgen in welke ondernemingen de Staat aandelen heeft (onderzoek [De staat als aandeelhouder](#)) hebben we jaarverslagen over het beheer van staatsdeelnemingen gebruikt. Deze informatie is aangevuld met dossieronderzoek bij onder meer de Kamer van Koophandel, beleidsnota’s en jaarverslagen van staatsdeelnemingen, juridische kennis van experts, gesprekken met aandeelhouders en ondernemingen en onderzoeken van derden naar staatsdeelnemingen.

4.3.4 Een feitelijk beschrijvende vraag

Feitelijk beschrijvende vragen worden vooral toegepast in focusonderzoek. Hierbij zijn document- en data-analyse belangrijke methoden.

- In het onderzoek [Voorarrest](#) is de vraag of Nederland Europees gezien koploper is op het gebied van voorarrest, beantwoord aan de hand van vier verschillende

internationale bronnen.

- In het onderzoek Toegang tot langdurige zorg zijn vooral kwantitatieve analysemethoden gebruikt om na te gaan in hoeverre afwijzingen voor Wlz-indicaties verklaard kunnen worden uit verschillen tussen gemeenten en tussen groepen aanvragers.

4.4 Conclusies

Let op: een beschrijvende vraag kan op zichzelf nooit leiden tot een conclusie! Een antwoord op beschrijvende onderzoeksvragen kan alleen in beschrijvende, neutrale termen worden gegeven, omdat je geen normen hebt gehanteerd. Voor een conclusie is nodig dat je bevindingen hebt: een vergelijking van de feitelijke situatie, die je beschreven hebt, met de norm die je gehanteerd hebt. Zie ook hoofdstuk 5 en hoofdstuk 10.

5.

Toetsende onderzoeksvragen

5.1 Inleiding

Nadat je de doel- en probleemstelling hebt geformuleerd, vertaal je dit naar een vraagstelling voor je onderzoek. De vraagstelling hangt af van het soort kennis waar je naar op zoek bent om je doelstelling te bereiken. Daarom is het belangrijk om goed na te denken over de formulering van je onderzoeksvragen. Uit een analyse van gepubliceerde rekenkamerrapporten blijkt dat in ons onderzoek in feite drie soorten vragen aan de orde komen: beschrijvende vragen, toetsende vragen en verklarende vragen. In de meeste onderzoeken wordt gebruik gemaakt van een combinatie van beschrijvende, toetsende en verklarende vragen. Toetsende vragen richten zich op kennis die je nodig hebt om uiteindelijk een oordeel te geven over het overheidsbeleid. We gaan nu in op het type 'toetsende onderzoeksvraag'.

5.2 Wat is een toetsende vraag?

Toetsende vragen maken een belangrijk onderdeel uit van rekenkameronderzoek. Toetsend onderzoek leidt in onze rapporten vrijwel altijd tot een bestuurlijke uitspraak: een oordeel over een bepaalde praktijk of situatie die de rekenkamer heeft onderzocht.

Let op: Als een toetsende onderzoeksvraag een kwalificatie bevat (bijvoorbeeld voldoende, goed, adequaat, toereikend), dan moet je daarvoor ook een norm formuleren. Anders zou je onderzoeksvraag direct leiden tot een oordeel. Als je een norm hebt geformuleerd, is de uiteindelijke, feitelijke conclusie: er is wel, niet of deels aan de norm voldaan. Op basis daarvan kan dan een oordeel worden geformuleerd. Zie ook 5.3.3 en paragraaf 10.3.

Kenmerkend voor toetsende vragen is dat ze leiden tot een feitelijke conclusie (de norm wordt in de praktijk wel of niet gehaald) en in veel gevallen leidt deze feitelijke conclusie tot een oordeel (wat vinden we ervan dat de norm wel of niet wordt gehaald). In hoofdstuk 10 wordt deze fase van het onderzoek in meer detail beschreven.

Binnen toetsend onderzoek kun je onderscheid maken naar verschillende vragen, van relatief eenvoudige vragen die geen causaliteitselement kennen (bijvoorbeeld: “Is de Tweede Kamer volledig geïnformeerd over onderwerp X?”) tot vragen die alleen beantwoord kunnen worden als we een controlegroep kunnen construeren. In hoofdstuk 7 vind je voorbeelden van toetsende vragen in relatie tot doelrealisatie, doelgroepbereik, doeltreffendheid en doelmatigheid.

5.3 Toetsen aan een norm

5.3.1 Inleiding

In veel rekenkameronderzoek wordt getoetst aan normen. Die toetsing leidt tot een (feitelijke) conclusie. Het is daarom belangrijk dat je expliciet aangeeft welke normen je in het onderzoek hanteert zodat duidelijk is waarop de conclusies van het onderzoek zijn gebaseerd. Let erop dat de (toetsende) onderzoeksvraag en de te hanteren normen goed op elkaar aansluiten.

Het kenmerkende van toetsend onderzoek is dat je norm (de gewenste situatie) en de praktijk (de werkelijke situatie) met elkaar vergelijkt. Als wij de wachttijden in de jeugdzorg onderzoeken zouden (norm)toetsende vragen in dit onderzoek kunnen zijn:

- Voldoen de wachttijden in de praktijk aan de door de minister daaraan gestelde normen?
- Is de Tweede Kamer juist en volledig geïnformeerd?

In dit onderzoek staat de norm zelf in principe niet ter discussie, maar de praktijk. Overigens: mocht uit het onderzoek blijken dat de door de minister gestelde normen – die ook onze normen waren – niet realistisch zijn, dan is het natuurlijk wel mogelijk een aanbeveling te doen waarin we de minister vragen de normen te heroverwegen of aan te passen.

Normen kun je relateren aan doelen, middelen, activiteiten, prestaties en effecten. In hoofdstuk 7 gaan we dieper in op de verschillende onderzoeksobjecten en -varianten.

5.3.2 Normen bepalen

Voor het beantwoorden van toetsende onderzoeksvragen heb je uiteraard normen nodig. Die normen kunnen afkomstig zijn uit verschillende bronnen: eigen normen

van de Algemene Rekenkamer en normen op basis van beleidsdocumenten en wetgeving. Soms bestaan er nog geen normen en zullen deze voor het onderzoek ontwikkeld moeten worden.

5.3.2.1 Normen Algemene Rekenkamer

De Algemene Rekenkamer heeft een groot aantal eigen normen. Typische normen die vaak in ons onderzoek terugkomen, zijn normen die gaan over de informatievoorziening aan het parlement, de kwaliteit van de beleidsvoorbereiding en de (beleids)informatie en de evalueerbaarheid van beleid. Zo stellen we in het onderzoek [Energie label voor koopwoningen](#) de toetsende onderzoeksvraag: "Is de Tweede Kamer in begroting, jaarverslag of andere Kamerstukken voldoende geïnformeerd door de betrokken bewindspersonen over de resultaten en de uitgaven?" (zie 5.3.3 voor meer concrete voorbeelden). Ook voor andere onderwerpen hebben we de afgelopen jaren op basis van diverse onderzoeken normenkaders in de vorm van basisnormen ontwikkeld. In hoofdstuk 8 vind je een uitgebreid overzicht van onderwerpen waarvoor normenkaders beschikbaar zijn. Voor sommige onderwerpen ontwikkelen we een normenkader tijdens of als resultaat van een onderzoek. Tijdens zo'n onderzoek is dat normenkader dan nog niet bruikbaar (tenzij je daar al overeenstemming over hebt met de gecontroleerde), maar achteraf is dat wel het geval. Enkele voorbeelden zijn de handreiking [Basisprincipes vastgoedmanagement](#) en de [Regeling grote projecten](#).

5.3.2.2 Bestaande bronnen

Naast bestaande rekenkamernormen zijn beleidsnotities, begrotingen en wet- en regelgeving belangrijke bronnen voor onze normen. Voorbeelden:

- Elk jaar onderzoeken we voor elk begrotingshoofdstuk de kwaliteit van de totstandkoming van de beleidsinformatie. De normen voor dit onderzoek komen uit de Comptabiliteitswet 2016: *beleidsinformatie is betrouwbaar en ordelijk tot stand gekomen en is niet strijdig met de financiële verantwoordingsinformatie*.
- Stel dat wij de kwaliteit van de rechtspraak willen meten, dan hebben we hiervoor normen nodig. Die zouden we kunnen ontleen aan de indicatoren die de Raad voor de Rechtspraak voor deze kwaliteit hanteert. Om de kwaliteit van rechtspraak te meten en monitoren, maakt de rechtspraak gebruik van indicatoren die betrekking hebben op verschillende kwaliteitsdimensies: juridisch-inhoudelijke kwaliteit, onpartijdigheid en integriteit, bejegening ter zitting en de doorlooptijd van behandelde zaken. Deze kwaliteitsindicatoren kunnen voor ons normen zijn. Overigens is het zelf meten van de juridisch-inhoudelijke kwaliteit niet simpel; we zouden daarvoor ook gebruik kunnen maken van de resultaten van het kwaliteitsstelsel van de rechtspraak, voor zover dat er is.

5.3.2.3 Nieuwe normen ontwikkelen

Lang niet altijd zijn er geschikte (externe) normen beschikbaar op het moment dat je bezig bent met het uitwerken van je onderzoeksplan. Dat betekent niet dat je met lege handen staat. Het betekent wel dat je zelf aan de slag moet met het zoeken naar geschikte normen of het zelf ontwikkelen van normen. Hierbij zou je ook gebruik kunnen maken van wetenschappelijke literatuur en best practice-voorbeelden. Het is verstandig altijd en in een vroeg stadium hierover met de gecontroleerde in gesprek te gaan en expliciet te vragen of hij hiermee instemt. De gecontroleerde hoeft het overigens niet altijd eens te zijn met de norm die wij hanteren. Maar het is wel belangrijk dat de gecontroleerde tijdig kennis heeft kunnen nemen van de norm die wij in het onderzoek gebruiken en dat wij weten welke argumenten hij heeft om het niet met ons eens te zijn. Dit voorkomt lastige discussies tijdens het onderzoek en bij de afstemming van het conceptrapport.

Let op: zorg dat je in je rapport, in de normenbijlage, altijd duidelijk maakt waarop de normen gebaseerd zijn.

5.3.2.4 Voorbeelden

Voor de beoordeling van de vraag of de minister voldoende beleidsprestaties en/of -effecten heeft geleverd is er in principe één algemene norm: de belofte die de minister in zijn beleidstukken gedaan heeft. Ieder beleid heeft eigen beoogde prestaties en effecten. Om deze reden zal je specifieke normen moeten afleiden uit de door de minister geformuleerde doelen en de gemaakte afspraken op het beleidsterrein in kwestie. Voorbeeld:

- In het vervolgonderzoek [Duurzaamheid intensieve veehouderij](#) is de vraag opgenomen "In hoeverre heeft de minister zijn beleidsdoelen wat het ammoniakbeleid en de bevordering van dierenwelzijn betreft gerealiseerd?". De normen voor het toetsen van deze vraag zijn de specifieke doelen voor het ammoniakbeleid en de bevordering van het dierenwelzijn zoals die in de beleidstukken zijn beschreven.
- In het onderzoek [Resultaten prijsonderhandelingen geneesmiddelen](#) was de vraag aan de orde of het Ministerie van VWS er in de prijsonderhandelingen met fabrikanten in slaagt het gestelde beleidsdoel (het bedingen van de adviesprijs van het Zorginstituut) te realiseren.

Voor de beoordeling van de vraag of de minister zijn prestaties en effecten voldoende doelmatig geleverd heeft zijn ook geen algemene normen beschikbaar. Doelmatigheid is een relatief begrip, wat betekent dat een uitspraak op de vraag of de doelmatigheid toereikend is, alleen gedaan kan worden door de verhouding tussen kosten en prestaties/effecten te vergelijken met een andere situatie.

Voorbeelden:

- In ons onderzoek [Bekostiging rechtspraak](#) (2016) is doelmatigheid gedefinieerd als een betere verhouding tussen de ingezette middelen en de geleverde prestaties. In de bijlage van het rapport geven we een toelichting op doelmatigheid in deze context: “alleen wanneer een productiviteitsstijging plaatsvindt met behoud van kwaliteit kan worden gesproken van een grotere doelmatigheid”. Vervolgens zijn productiviteit en kwaliteit verder gedefinieerd.
- In ons onderzoek [Fiscale stimulering van elektrische auto's](#) (2019) bespreken we de vraag “of het bedrag van de fiscale stimulering niet te hoog is in relatie tot de vermindering van de CO₂-uitstoot, zowel per auto als in totaal.” Ook in onze verantwoordingsonderzoeken over 2013 en 2014 is deze vraag gesteld. We hebben toen geconcludeerd dat de fiscale stimulering van zuinige auto's, gegeven de door het Ministerie van Financiën berekende belastingderving van € 1.000 per vermeden ton CO₂, relatief duur is en dat het kabinet vermoedelijk meer CO₂ had kunnen besparen bij een andere besteding van dat geld. Hier is dus de vergelijkingsmaatstaf (de norm) de kosten per vermeden ton CO₂.

5.3.3 De onderzoeksvraag in een normatief onderzoek

De onderzoeksvraag in een normatief onderzoek is van het type:

- Voldoet de beleidsvoorbereiding van beleid X aan de normen Y?
- Voldoen de beleidsdoelen voor beleid Y aan de norm Z?
- Voldoet de uitvoering van beleid X aan norm P?
- Voert minister A beleid Y uit zoals door hem bedoeld?
- Informeert minister B de Tweede Kamer over beleid X voldoende?

Bedenk hierbij dat het antwoord op de onderzoeksvraag meestal niet ‘ja’ of ‘nee’ is. Je kunt daar rekening mee houden door de vraag te beginnen met “In hoeverre...”.

Voorbeelden van dit soort vragen uit onze eerdere onderzoeken zijn:

- Is de Tweede Kamer in begroting, jaarverslag of andere Kamerstukken voldoende geïnformeerd door de betrokken bewindspersonen over de resultaten en de uitgaven? (onderzoek [Fiscale tegemoetkoming voor experts uit het buitenland](#)).

Hierbij werd de volgende (basis)norm gehanteerd:

“De verantwoordelijke bewindspersoon moet de Tweede Kamer voldoende informeren: relevante en getrouwe weergave (juist (overeenkomend met bron; niet gekleurd), volledig en tijdig), begrijpelijk en vergelijkbaar (voor zover relevant, bijvoorbeeld in de tijd), betrouwbaar, valide.”

- Is er sprake van een goede beleidsvoorbereiding door de minister? (onderzoek [Inburgering](#)). De norm is ‘goede beleidsvoorbereiding’, die als volgt wordt geconcretiseerd: “De doelen kunnen gerealiseerd worden met de beschikbare mensen en middelen. We verwachten dat de minister van SZW in

overleg met relevante organisaties bij de totstandkoming is nagegaan of het beleid uitvoerbaar (ook in tijd) en handhaafbaar is en ex-ante is nagegaan of de maatregelen de beoogde werking zullen hebben.”

- Heeft het instemmingsrecht van de medezeggenschap op de hoofdlijnen van de begroting in de periode 2015–2017 voor de voorinvesteringen gewerkt zoals beoogd? (onderzoek [Voorinvestering en medezeggenschap hoger onderwijs](#)). De norm is “het beleid heeft gewerkt zoals beoogd”. Ook hiervoor hebben we geen eigen norm. Je zult dus zelf op zoek moeten gaan naar een uitwerking. In het onderzoek is de norm als volgt geoperationaliseerd:

“In ons onderzoek zien we een besteding als een voorinvestering indien: 1) er een besluit genomen is om iets extra’s te doen (voorheen nog niet begroot); of 2) deze voorheen voor een later jaar stond ingepland en naar voren is gehaald; of 3) er meer geld wordt uitgetrokken (extra besteding) dan eerder was begroot; en 4) het geen gesubsidieerde besteding is, maar een besteding uit eigen middelen; en 5) de besteding niet ten koste is gegaan van lopende plannen en uitgaven op het gebied van kwaliteit; en 6) de besteding (ook) bijdraagt aan een verbetering van de kwaliteit van het onderwijs in 2015, 2016 en/of 2017, dat wil zeggen voor de huidige studenten.”

5.4 Methoden: hoe zijn toetsende vragen in eerder onderzoek beantwoord?

In hoofdstuk 9 gaan we uitgebreider in op methoden en technieken. Hier noemen we bij wijze van illustratie een aantal voorbeelden van gebruikte methoden in een toetsend onderzoek.

5.4.1 Casusonderzoek

Een voorbeeld van een casestudy-aanpak bij een toetsende vraag is het onderzoek [Vastgoed bij universiteiten](#). Voor de beantwoording van de vragen is onderzoek uitgevoerd bij zes universiteiten. De conclusie die wordt getrokken gaat uitsluitend over de zes onderzochte universiteiten.

5.4.2 Informatievoorziening

Er zijn ook toetsende vragen die geen betrekking hebben op een populatie maar die gaan over de vraag of een partij (meestal de betreffende minister of de Tweede Kamer) voldoende is geïnformeerd over een bepaald onderwerp. Hier wordt de (kwaliteit van de) informatie die de Tweede Kamer ontvangt, onderzocht. Beoordeeld wordt of deze voldoende is, aan de hand van de Algemene Rekenkamernorm. Een voorbeeld van een onderzoek waar een dergelijke vraag wordt gesteld is het

onderzoek [Fiscale tegemoetkoming voor experts uit het buitenland](#). Hiervoor is de informatie in de documenten die de Tweede Kamer jaarlijks ontvangt (de Miljoenennota, de begroting en het jaarverslag) onderzocht.

5.5 Conclusies en oordelen

Met een bevinding geef je aan of en in hoeverre de praktijk afwijkt van de norm. Een conclusie is het (feitelijke) antwoord op een onderzoeksvraag en is opgebouwd uit een of meer bevindingen. Een conclusie kan zijn: “De eindafrekening kwam binnen na de afgesproken deadline” (de norm). Het bestuurlijke oordeel dat daarbij hoort, geeft aan of wij het erg vinden en zou kunnen luiden: “Wij zien het als een ernstige tekortkoming dat de eindafrekening te laat binnenkwam....” Een dergelijke uitspraak is een kwalificatie, een waardeoordeel. Voor deze kwalificatie hebben we geen normen; nergens staat wanneer we een afwijking van de norm erg moeten vinden. Het betreft een bestuurlijke afweging en keuze. In hoofdstuk 10 wordt uitvoeriger ingegaan op dit onderwerp.

6.

Verklarende onderzoeksvragen

6.1 Inleiding

Nadat je de doel- en probleemstelling hebt geformuleerd, vertaal je dit naar een vraagstelling voor je onderzoek. De vraagstelling hangt af van het soort kennis waar je naar op zoek bent om je doelstelling te bereiken. Daarom is het belangrijk om goed na te denken over de formulering van je onderzoeksvragen. Uit een analyse van gepubliceerde rekenkamerrapporten blijkt dat in ons onderzoek in feite drie soorten vragen aan de orde komen: beschrijvende vragen, toetsende vragen en verklarende vragen. In de meeste onderzoeken wordt gebruik gemaakt van een combinatie van beschrijvende, toetsende en verklarende vragen. Verklarende vragen richten zich op kennis die je nodig hebt om mogelijke verklaringen of oorzaken te vinden voor bepaalde situaties of voor geconstateerde problemen en tekortkomingen. We gaan nu in op het type 'verklarende onderzoeksvraag'.

6.2 Wat is een verklarende vraag?

6.2.1. Inleiding

Het doel van verklarende vragen is om mogelijke verklaringen te vinden voor bepaalde situaties of voor geconstateerde problemen en tekortkomingen. Bijvoorbeeld als we vinden dat de verklaringen die anderen tot nu toe hebben gegeven voor een bepaalde ontwikkeling of situatie onbevredigend zijn of als we meer willen weten over eventuele oorzaken van bepaalde ontwikkelingen of problemen. In wetenschappelijke termen zouden we spreken van *hypothesen*, maar we gebruiken hier een meer praktische terminologie. Dat doen we ook om het misverstand te vermijden dat rekenkamers actief participeren in het

wetenschappelijk discours. Dat laat uiteraard onverlet dat ook rekenkameronderzoek moet voldoen aan wetenschappelijke standaarden.

Een verklarende onderzoeksvraag is van het type “Wat zijn (mogelijke) verklaringen voor ...?”. Deze vraag is ook van belang als je verklaringen zoekt voor het verschil tussen de norm en de werkelijkheid die je aantreft (zie hoofdstuk 5).

In hoofdstuk 7 vind je voorbeelden van verklarende vragen in relatie tot doelrealisatie, doelgroepbereik, doeltreffendheid en doelmatigheid.

In deze handleiding maken we onderscheid tussen verklarende vragen die erop gericht zijn om mogelijke (indicatieve) verklaringen voor een probleem te vinden (6.2.2) en verklarende vragen die het doel hebben om mogelijke verklaringen ook daadwerkelijk te valideren (6.2.3). In 6.2.4 vind je een uitgebreid overzicht van mogelijke oorzaken of verklaringen voor geconstateerde problemen of tekortkomingen. Het is belangrijk om zorgvuldig na te gaan in hoeverre de minister of het bestuursorgaan verantwoordelijk kan worden gehouden voor het probleem en in hoeverre de minister of het bestuursorgaan voldoende bevoegd en in staat is om het probleem op te lossen.

6.2.2 Een verklarende vraag gericht op indicaties

6.2.2.1 Inleiding

Het is niet altijd nodig of mogelijk (binnen de gestelde termijn) om een gevalideerde verklaring te geven voor tekortkomingen, achterblijvende prestaties en effecten of ondoelmatigheden. Toch wil je de toegevoegde waarde van je onderzoek vergroten door indicaties te geven voor oorzaken van geconstateerde problemen of tekortkomingen. Je kunt in dat geval kiezen voor het onderzoeken van een of enkele casussen (6.2.2.2) of het inventariseren van mogelijke verklaringen aan de hand van bijvoorbeeld literatuurstudie en het interviewen van experts (6.2.2.3).

6.2.2.2 Casusonderzoek

Bij het analyseren van mogelijke verklaringen voor problemen of tekortkomingen kun je gebruik maken van casusonderzoek. In een casusonderzoek selecteer je de casussen voor je onderzoek doorgaans op basis van informatierijkheid (in dit geval casussen die inzicht kunnen geven in mogelijke verklaringen). De casussen worden dus niet gekozen omdat ze representatief zijn voor een grotere populatie.

We illustreren dit met het voorbeeld van een onderzoek naar oorzaken van de kostenvermindering van windparken op zee. We vinden de verklaringen die anderen tot nu toe hebben gegeven van onvoldoende kwaliteit (bijvoorbeeld slecht onderbouwd), of we kunnen niet voortbouwen op inzichten van anderen omdat die er niet zijn. Maar we willen wel meer te weten komen over eventuele oorzaken van

kostenvermindering van windparken op zee. We onderzoeken een aantal gevallen om de mogelijke oorzaken te verkennen. Op basis hiervan formuleren we mogelijke verklaringen voor de kostenvermindering van windparken op zee. Deze verklaringen zijn de uitkomsten van ons onderzoek en zijn daarmee (uiteraard) van toepassing op de door ons onderzochte gevallen. Om de vraag te beantwoorden of ze ook opgaan in andere gevallen is aanvullend onderzoek nodig.

Let op dat je in je rapport de beperkingen van je analyse duidelijk markeert.

6.2.2.3 Inventarisatie

Je kunt mogelijke verklaringen voor problemen of tekortkomingen ook proberen te achterhalen door bijvoorbeeld literatuurstudie, (internationale) vergelijkingen, expertmeetings en observaties.

We illustreren dit met het voorbeeld van een onderzoek naar oorzaken van de kostenvermindering van windparken op zee. We hebben ons bij de opzet en uitvoering van ons onderzoek naar kostenvermindering van windparken op zee niet gericht op het zoeken of vinden van oorzaken. Maar nu we onze bevindingen en conclusies over kostenvermindering hebben geformuleerd, willen we weten welke factoren van invloed zouden kunnen zijn op de kostenvermindering. We vermoeden dat de vermindering veroorzaakt zou kunnen worden door een aantal factoren die we nader omschrijven. Inspiratiebronnen daarvoor zijn bijvoorbeeld de verklaringen die door de onderzochte instantie zelf of door andere informanten in ons onderzoek zijn genoemd. We formuleren hier mogelijke oorzaken als verklaring, maar niet op basis van onderzoek dat daarop gericht is.

Let op dat je in je rapport de beperkingen van je analyse duidelijk markeert.

6.2.3 Een verklarende vraag gericht op validering

6.2.3.1 Inleiding

Aanleiding voor rekenkameronderzoek kan zijn dat we willen nagaan welke oorzaken zijn aan te wijzen voor tekortkomingen, achterblijvende prestaties en effecten of ondoelmatigheden. Je kunt ervoor kiezen om voor het zoeken van verklaringen gebruik te maken van onderzoek van derden (6.2.3.2), zelf onderzoek te verrichten (6.2.3.3), of een combinatie daarvan. Daarnaast kun je nagaan of verklaringen te vinden zijn in eerder rekenkameronderzoek.

6.2.3.2 Onderzoek van derden naar verklaringen

Verklaringen hoef je niet altijd zelf te onderzoeken. Je kunt ook nagaan of je hiervoor onderzoek van derden kunt gebruiken. Als je de uitkomsten van die onderzoeken wilt gebruiken, is het noodzakelijk dat je de kwaliteit daarvan beoordeelt. Het onderzoek

moet bijvoorbeeld op een zorgvuldige en controleerbare wijze zijn uitgevoerd, de gebruikte methoden moeten betrouwbaar en nauwkeurig zijn en het onderzoek moet voldoende onafhankelijk zijn uitgevoerd. In hoofdstuk 08 wordt verder ingegaan op het beoordelen van onderzoek van derden. Als je geconstateerd hebt dat de kwaliteit van het onderzoek voldoende is, kun je de uitkomsten gebruiken in je rapport. Het is belangrijk dat je altijd de bron van die uitkomsten in je rapport vermeldt.

We illustreren dit aan de hand van een voorbeeld: een onderzoek naar oorzaken van de kostenvermindering van windparken op zee. We kunnen nagaan of anderen kostenvermindering van windparken op zee hebben onderzocht en welke oorzaken ze daarvoor hebben gevonden. Als we die verklaringen overtuigend vinden, hun onderzoek op een valide en betrouwbare manier is uitgevoerd, en de resultaten daarvan ook te generaliseren zijn naar de gevallen van 'windparken op zee' die in ons eigen onderzoek centraal staan, kunnen we de conclusies overnemen in ons eigen rapport. We beschrijven dan de onderzoeksresultaten van anderen. Als we het hierbij houden, hebben we zelf geen mogelijke verklaringen geformuleerd en deze ook niet zelf getoetst in gevallen die we zelf onderzocht hebben. We hebben (alleen) verklaringen overgenomen van anderen en hebben geconstateerd dat die 'waar' zijn.

6.2.3.3 Eigen onderzoek naar verklaringen

Als we geen gebruik kunnen maken van onderzoeken van derden (omdat ze er niet zijn of omdat de kwaliteit ervan onvoldoende is), dan kunnen we zelf op zoek gaan naar verklaringen. Daarbij probeer je op basis van bijvoorbeeld literatuurstudie en gesprekken met deskundigen veronderstellingen te formuleren die het fenomeen kunnen verklaren. Houd er in de planning van je onderzoek rekening mee dat dit een arbeidsintensief proces kan zijn.

6.2.4 Analyse van oorzaken

6.2.4.1 Interne en externe oorzaken

Het is belangrijk om inzicht te hebben in oorzaken van tekortkomingen, achterblijvende prestaties en effecten of ondoelmatigheden. Hiermee kun je namelijk relevante aanbevelingen doen aan de ministers, bestuursorganen en/of de Tweede Kamer en voorkom je dat aanbevelingen 'omgekeerde conclusies' zijn. De oorzaken voor achterblijvende prestaties en effecten of ondoelmatigheden kunnen zowel intern (endogeen) als extern (exogeen) van aard zijn.

- *Interne oorzaken* zijn oorzaken die liggen binnen de invloedssfeer van de minister. Hierbij kun je denken aan: onvoldoende duidelijke of realistische formulering van het doel van het beleid, het inzetten van de verkeerde beleidsinstrumenten, het ondernemen van de verkeerde activiteiten en het creëren van onvoldoende

randvoorwaarden voor doeltreffendheid. Maar ook kenmerken van de organisatie, zoals de (kwaliteit van de) personele bezetting en cultuur en gedrag vallen hieronder.

- *Externe oorzaken* liggen buiten de invloedssfeer van de minister. De oorzaken van achterblijvende prestaties en effecten of ondoelmatigheden zijn niet of nauwelijks door de organisatie te beïnvloeden. Voorbeelden van externe oorzaken zijn: onvoldoende steun voor het gevoerde beleid, bestuurlijke complexiteit (denk aan het complexe gebouw van de sociale zekerheid, functioneel en territoriaal gedecentraliseerd beleid, marktwerking en privatisering), onvoldoende mogelijkheden tot effectvergroting, andere (beleids)ontwikkelingen die het mogelijke positieve effect van het gevoerde beleid tenietdoen, en conjuncturele ontwikkelingen.

Overigens mag je wel verwachten dat de minister zo goed mogelijk geanticipeerd heeft op externe invloeden.

6.2.4.2 De verklaringenmatrix

Om het brede palet aan mogelijke verklaringen te schetsen gebruiken we een matrix (zie tabel 6.1). Op de verticale as wordt een onderscheid gemaakt naar de beïnvloedingsmogelijkheden van de minister. In feite gaat het hier om de ministeriële verantwoordelijkheid. Op de horizontale as wordt een onderscheid gemaakt tussen mogelijke verklaringen voor achterblijvende prestaties en voor achterblijvende effecten. Die prestaties hebben we ruim opgevat. Voor de burger maakt het immers niet uit of een minister of een uitvoerende instantie verantwoordelijk is.

Let op: met deze matrix willen we je op weg helpen bij het vinden van verklaringen; het is geen uitputtende opsomming. Ga er bij je eigen analyse ook niet zomaar vanuit dat verklaringen (alleen) op het proces betrekking hebben.

Tabel 6.1 De verklaringenmatrix

	Mogelijke verklaringen voor achterblijvende prestaties van het Rijk en daarmee verbonden organen én lokale overheden (in medebewind)	Mogelijke verklaringen voor achterblijvende maatschappelijke effecten
Binnen directe beïnvloedingssfeer minister	Organisatie en uitvoering van beleid <ul style="list-style-type: none"> • Mismatch tussen prestatiedoelen, tijd, mensen en middelen (te veel of te weinig ambitieus) • Doelen niet SMART-C • Strijdig beleid van andere ministeries • Gebrekkig wetgevend kader • Gebrekkige handhaving • Bedrijfsvoeringsproblemen: onvoldoende financieel en materieelbeheer, ontoereikende ICT, onrechtmatigheden, gebrekkige beleidscontrolfunctie • Organisatiecultuur ministerie, kwaliteit personeel, ziekteverzuim, kwaliteit werkomgeving, gedrag • Onduidelijke taken en bevoegdheden • Discongruente gebiedsindelingen 	Kwaliteit van beleid en uitvoering <ul style="list-style-type: none"> • Gebrekkige beleidstheorie (instrumentkeuze is ontoereikend of geen rekening gehouden met neveneffecten) • Onvoldoende adaptief vermogen om in te spelen op veranderende omstandigheden • Achterblijvende prestaties van het Rijk
Buiten directe beïnvloedingssfeer minister	Betrokkenheid van en uitvoering door derden <ul style="list-style-type: none"> • Ontbreken van draagvlak bij stakeholders (deels binnen en deels buiten de beïnvloedingssfeer; dit moet per geval worden beoordeeld) • Strijdig beleid van EU, provincies en gemeenten • Uitvoerders die wel willen, maar niet kunnen 	Overmachtssituaties <ul style="list-style-type: none"> • Ingrijpende gebeurtenissen, zoals oorlog, ongelukken en natuurrampen • Economische, demografische of klimatologische ontwikkelingen • Strijdig beleid van EU, provincies en gemeenten

De vier categorieën in de matrix sluiten elkaar niet volledig uit. Er blijft sprake van een grijs gebied, bijvoorbeeld rond het begrip draagvlak (is de minister daarvoor verantwoordelijk, of liggen de belangen van diverse betrokkenen zo ver uit elkaar dat het niet realistisch is om te stellen dat de minister hierop direct invloed kan uitoefenen).

Het is belangrijk om te blijven zoeken naar achterliggende verklaringen voor de gevonden verklaringen. Je kunt je bijvoorbeeld afvragen waarom is er sprake van een gebrekkig wetgevend kader en constateren dat de juridische expertise op het ministerie tekortschiet. Vervolgens kun je je afvragen waarom dit het geval is en constateren dat geen prioriteit is gegeven aan het werven van gekwalificeerde juristen of dat geen gebruik is gemaakt van de wetgevingsjuristen op het ministerie van Justitie en Veiligheid. Streef ernaar om uiteindelijk inzicht te krijgen in de *root cause* van de problemen en tekortkomingen. Dan kan het probleem ook bij de kern worden aangepakt.

6.3 Methoden: hoe zijn verklarende vragen in eerder onderzoek beantwoord?

In hoofdstuk 9 gaan we uitgebreider in op methoden en technieken. Hier noemen we bij wijze van illustratie een aantal voorbeelden van gebruikte methoden in een verklarend onderzoek.

Je kunt bij een verklarend onderzoek verschillende methoden en technieken gebruiken, bijvoorbeeld door de situatie mét de vermeende oorzaak te vergelijken met de situatie zónder de vermeende oorzaak. Daarbij kun je verschillende gevallen (casussen) met elkaar vergelijken op hetzelfde moment, of één casus vergelijken in de tijd.

Voorbeelden van verklarende vragen uit eerdere onderzoeken zijn:

- Wat zijn mogelijke verklaringen voor het feit dat het niet gelukt is een evenwicht tussen bekostiging en uitgaven te realiseren? (onderzoek [Passend-onderwijs](#)).
- Waarom stijgt het aantal schadevergoedingen? (onderzoek [Voorarrest](#)).
- Wat zijn de oorzaken van de kostenvermindering? (onderzoek [Kosten windenergie op zee](#)).

6.4 Soorten conclusies

Een verklarende vraag kan leiden tot een conclusie, maar dat hoeft niet. Je kunt alleen conclusies trekken op basis van bevindingen: een vergelijking van de feitelijke situatie die je hebt aangetroffen met de norm die je gehanteerd hebt. Zie ook hoofdstuk 5, paragraaf 10.2 en 10.4.

Het is daarom belangrijk dat je bij het trekken van conclusies goed voor ogen houdt wat je onderzoeksvraag is en op welke manier je je casussen hebt geselecteerd.

Welk type uitspraak kun je daarmee onderbouwen, en ook: welke niet? In paragraaf 10.3 en 10.5 gaan we uitvoeriger in op dit onderwerp.

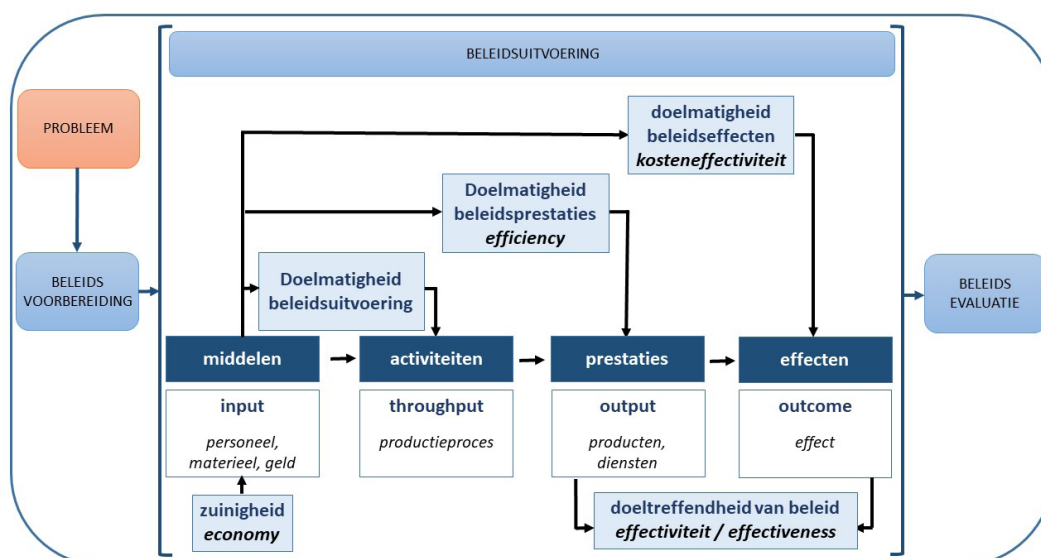
7.

Onderzoeksubjecten en -varianten

7.1 Inleiding

Je kunt het functioneren van de publieke sector tot op zekere hoogte vergelijken met een productieproces. In dat proces onderscheiden we vier aspecten: middelen, activiteiten, prestaties en effecten, ofwel input, throughput, output en outcome. Door je onderzoek te richten op verschillende combinaties van deze aspecten kun je uitspraken doen over doelrealisatie, doeltreffendheid en doelmatigheid van beleid of uitvoering. Bij doelrealisatie en doeltreffendheid ga je uit van de doelstellingen van het beleid.

Figuur 7.1 Productieproces publieke sector



Het is niet altijd mogelijk om duidelijk onderscheid te maken tussen prestaties en effecten. Vaak is er sprake van een keten van prestaties en effecten, waarbij het een kwestie van afspraak is waar de grens wordt getrokken. Als je dit in het onderzoek moet doen, zorg dan in ieder geval dat je er met de 'gecontroleerde' duidelijkheid over krijgt waar de bewindspersoon of het bestuur van een instelling nog verantwoordelijk voor is en waar niet meer voor.

Voorbeeld: het UWV heeft onder meer als taak (doel van het beleid) werklozen met een afstand tot de arbeidsmarkt aan een baan te helpen. Hiertoe biedt het UWV werkzoekenden vacatures aan. Is deze aanbieding op zichzelf een prestatie of is het pas een prestatie als het aanbod tot een sollicitatie of tot een plaatsing (werk) heeft geleid? Of is dit laatste een effect? En wanneer telt een plaatsing mee? Als de werkzoekende de baan accepteert of pas als hij een bepaalde minimale periode is blijven werken?

7.2 Middelen

Bij middelen gaat het om alle inzet die nodig is geweest om de effecten dan wel prestaties te realiseren: het ingezette personeel, het verbruikte materiaal, de ingehuurd diensten en de uitgaven aan (niet-) financiële beleidsinstrumenten, zoals subsidies, leningen en voorlichtingsprogramma's.

Om de totale inzet van middelen te bepalen is het voor het meten van prestaties en effecten noodzakelijk de verschillende middelen in één grootte uit te drukken. De meest voor de hand liggende is geld. Het uitdrukken van middelen in geld kan op twee manieren: in termen van uitgaven en in termen van kosten. Uitgaven zijn daadwerkelijk in een periode betaalde bedragen. Kosten (of lasten) zijn (potentiële) uitgaven die aan een bepaalde periode worden toegerekend. Hierbij is niet het moment van betaling, maar het verbruiksmoment bepalend. Het is belangrijk dat je deze begrippen correct gebruikt..

7.3 Activiteiten

Onder 'activiteiten' verstaan we al datgene wat bij een ministerie of bij een derde wordt uitgevoerd om middelen te kunnen omzetten in prestaties voor het beleid. Met andere woorden: activiteiten hebben betrekking op de uitvoering van het beleid – niet op de uitkomsten van het beleid (dat zijn immers prestaties en/of effecten). Voorbeelden van activiteiten zijn het beoordelen van vergunningen, het werven en opleiden van personeel, het opzetten van een geautomatiseerd systeem en het verzamelen van relevante beleidsinformatie. Het zijn altijd handelingen.

Het kan van belang zijn om de activiteiten mee te nemen in onderzoek naar prestaties en effecten. Verkeerd gekozen activiteiten kunnen een verklaring bieden voor onder andere achterblijvende prestaties en/of effecten (zie ook hoofdstuk 06).

7.4 Prestaties

Prestaties zijn de resultaten van werkprocessen die binnen een organisatie worden doorlopen om een beleidsdoelstelling te realiseren. We onderscheiden (beleids) prestaties van het Rijk en prestaties van derden.

7.4.1 Prestaties van het Rijk

Prestaties van het Rijk zijn de aan de doelstelling van het beleid gerelateerde resultaten van werkprocessen van een organisatie (onderdeel) van het Rijk die bestemd zijn voor de 'buitenwereld'. Prestaties van het Rijk kunnen verschillende soorten producten en diensten behelzen. Enerzijds gaat het om prestaties die extern merkbaar zijn bij burgers, maatschappelijke instellingen en bedrijven. Voorbeelden van zulke prestaties zijn subsidies, vergunningen en voorlichtingscampagnes. Anderzijds zijn er 'faciliterende' prestaties waarbij de ministeries een concrete dienst verlenen aan derden, zoals gemeenten of rwt's/zbo's. Voorbeelden hiervan zijn: wet- en regelgeving, een covenant of contract, een geautomatiseerd systeem of de verzamelde relevante beleidsinformatie.

Sommige van deze prestaties zijn meer intern dan extern gericht en lijken daarmee eerder onderwerp voor een beheeronderzoek. Toch kan dit soort prestaties, bijvoorbeeld een geautomatiseerd systeem voor registratie van mestoverschotten, wel eens het enige zijn wat je op een bepaald beleidsterrein mag verwachten van de rijksoverheid. Daarmee is het dus ook een relevante prestatie voor het bereiken van een extern effect. Faciliterende prestaties komen vooral voor bij territoriaal of functioneel gedecentraliseerd beleid. Zie hiervoor ook paragraaf 7.3.

7.4.2 Prestaties van derden

In veel gevallen wordt beleid van de rijksoverheid uitgevoerd door derden, zoals rwt's, zbo's, gemeenten en provincies. De uitvoering van het beleid door derden impliceert dat zij ook verantwoordelijk zijn voor het leveren van de 'echte' extern gerichte prestaties, de finale prestaties aan burgers, maatschappelijke instellingen en bedrijven. Prestaties van derden kun je omschrijven als "de aan de doelstelling van het rijksbeleid gerelateerde resultaten van werkprocessen van een organisatie- (onderdeel) buiten het Rijk die bestemd zijn voor burgers, maatschappelijke instellingen en bedrijven". Een voorbeeld: bij het beleid gericht op het verminderen van de hoeveelheid bedrijfsafval in Nederland betreffen de beleidsprestaties van het

Rijk onder meer relevante wet- en regelgeving en een faciliteit voor de vergunningverlening van lokale overheden aan bedrijven. In dit voorbeeld is het uiteindelijke effect, te weten minder afval, afhankelijk van onder andere de verleende vergunningen door lokale overheden (de prestatie). Andere voorbeelden van prestaties van derden zijn: het aantal nieuwe politieagenten (te leveren prestatie voor veiligheidsbeleid) en het aantal beschikbare plaatsen voor inburgeringcursussen (te leveren prestatie voor inburgeringsbeleid). Zie verder ook paragraaf 7.3.

Vaak is de gemeente of provincie ook nog een doorgeefluik voor uitvoerende organisaties. Het Rijk geeft geld dat vervolgens weer verdeeld wordt over instellingen. De instellingen leveren de uiteindelijke prestaties.

De prestaties van rwt's kunnen ook zelfstandig object van rekenkameronderzoek zijn. Voor het onderzoeken van de prestaties van rwt's gelden in principe dezelfde regels als voor onderzoek naar prestaties van het Rijk. Bijzonder aan het rwt-onderzoek is wel – in vergelijking met andere derden die prestaties leveren voor rijksbeleid – dat ook de bedrijfsvoering van een rwt kan worden onderzocht.

7.5 Effecten

Effecten zijn de gevolgen van beleid van het Rijk of een rwt die merkbaar zijn in de maatschappij. Voorbeelden van (beleids)effecten zijn: de reductie van CO₂-uitstoot (beoogd effect van klimaatbeleid), meer mensen die zich veilig voelen op straat (beoogd effect van veiligheidsbeleid) en betere inburgering van migranten (beoogd effect van inburgeringsbeleid).

7.5.1 Beoogde effecten

Het onderzoek naar effecten is gecompliceerd, omdat het lastig is deze te kwantificeren (bijvoorbeeld: 'goede volksgezondheid in Nederland') en omdat vaak moeilijk is vast te stellen in hoeverre deze effecten door het beleid zijn gerealiseerd. Bij effecten kunnen immers behalve de minister (of andere bij het beleid betrokken partijen) ook andere factoren van invloed zijn, zoals demografische ontwikkelingen of de situatie op de arbeidsmarkt. De rijksoverheid wil bijvoorbeeld dat het UWV mensen aan het werk helpt. De mogelijkheid om mensen aan een baan te helpen wordt niet alleen beïnvloed door de activiteiten van het UWV, maar ook door bijvoorbeeld de situatie op de arbeidsmarkt (het aantal beschikbare vacatures) en daarmee samenhangend de economie (meer aanbod van werk in een periode van hoogconjunctuur).

Ook de lengte van de 'beleidsketen' kan complicerend werken. Tussen directe overheidsprestaties en de uiteindelijk gerealiseerde maatschappelijke effecten zitten vaak vele schakels, zoals bij subsidies voor milieubeleid: uitgekeerde subsidies,

bekendheid doelgroep met subsidie, aanschaf product door doelgroep, gebruik van product door doelgroep, veranderingen in milieu. Hoe verder een prestatie of effect in de keten gelegen is, des te moeilijker de directe causale relatie is te leggen met ingezet beleid.

Om de effectiviteit van het beleid te kunnen bepalen, moet de invloed van externe factoren worden geëlimineerd. Dat is een lastige klus die veel tijd kost en waarvan de uitkomsten onzeker zijn. Daarom kun je soms beter volstaan met het onderzoeken van prestaties of, als dit mogelijk is, het onderzoeken van de effecten van de geleverde prestaties. Immers, geleverde prestaties leiden niet noodzakelijkerwijs tot het beoogde (maatschappelijke) effect. Ze zijn hoogstens een noodzakelijke, maar geen voldoende voorwaarde voor het realiseren van het beoogde eindeffect.

Tegelijk is een oordeel over effecten belangrijk omdat dit de relevantie van het onderzoek verhoogt. Voor de burger telt uiteindelijk vaak niet de beleidsprestatie, maar de mate waarin het (maatschappelijk) probleem wordt opgelost of verminderd. De minister is daarop aanspreekbaar.

Een andere reden om effecten te onderzoeken is het gevaar dat de toegenomen aandacht voor sturing op basis van prestaties met zich kan meebrengen dat je de inhoudelijke doelstelling uit het oog verliest (het optreden van de performance paradox). Zo kan, als het onderwijs wordt bekostigd op basis van het aantal afgestudeerden, de neiging ontstaan in te leveren op het niveau van de afgestudeerden. Zie ook 7.2.5.2. Bij onderzoek naar de hele keten zou dit wel kunnen worden opgemerkt, bijvoorbeeld doordat de betrokken generatie studenten minder snel werk kan vinden (waarbij je wel veronderstelt dat de overige omstandigheden, zoals de arbeidsmarktsituatie, gelijk blijven).

7.5.2 Neveneffecten

Het kan van belang zijn om in een onderzoek naar prestaties en effecten ook eventuele (negatieve én positieve) neveneffecten te bekijken. Vooral bij ongewenste neveneffecten is dat relevant.

Een voorbeeld van een positief neveneffect is een lockdownmaatregel om de verspreiding van een virus af te remmen met als positief neveneffect dat de luchtkwaliteit in steden aanmerkelijk verbetert.

Voorbeelden van ongewenste neveneffecten zijn perverse prikkels (een hogeschool wordt gefinancierd op basis van het aantal gediplomeerden, wat kan leiden tot verlaging van examennormen en verlies van kwaliteit) en het waterbedeffect (effectieve bestrijding drugscriminaliteit leidt tot toename van andersoortige criminaliteit).

Onderzoek naar neveneffecten wordt gecompliceerd door het feit dat dit in principe alle kanten op kan; het wordt niet begrensd door beleidsdoelen. Een manier om toch

een afbakening aan te brengen is: focussen op neveneffecten die in ander dan het onderzochte beleid juist worden bestreden (bijvoorbeeld ongunstige milieueffecten van economisch stimuleringsbeleid).

Onbedoelde effecten van beleid kunnen een indicatie zijn dat de onderliggende beleidstheorie niet goed is uitgewerkt. Naast de bedoelde effecten kan ook ander gedrag worden aangemoedigd waar niet in was voorzien en dat het bedoelde beleidseffect beperkt of zelfs tenietdoet. Een grondige ex ante-evaluatie is een manier om ongewenste neveneffecten zo veel mogelijk in de basis te voorkomen: nauwgezet theoretisch doorexerceren wat de mogelijke (onbedoelde) gevolgen van het beleid of de interventie zijn. Zie ook hoofdstuk 8.

7.6 Onderzoek naar prestaties en effecten van derden bij gedecentraliseerd beleid

- Als sprake is van gedecentraliseerde uitvoering van beleid moet je eerst vaststellen waarvoor de minister precies verantwoordelijk is. Het onderzoek dient aan te sluiten bij de verantwoordelijkheid van de minister. Vervolgens moet je vaststellen wat de minister heeft gedaan (dat wil zeggen welke prestaties hij/zij heeft geleverd) om ervoor te zorgen dat het beleid op decentraal niveau goed kan worden uitgevoerd. Daarbij ga je ook na of de minister voldoende beleidsinformatie heeft verzameld over deze 'rijksprestaties'.
- Als beleid gedecentraliseerd wordt uitgevoerd, ga je na in hoeverre de minister beschikt over deugdelijke informatie over de voortgang van de uitvoering van het beleid (dus ook over prestaties en effecten). Als de minister niet beschikt over dergelijke informatie, kan de rekenkamer besluiten zelf die informatie te verzamelen. Op die manier kan zij onderbouwde conclusies trekken over de mate waarin de prestaties/effecten van derden worden behaald of de mate waarin het beleid effectief is.
- Een verklaring voor het niet realiseren van prestaties door bijvoorbeeld gemeenten en van effecten in de maatschappij kan zijn dat het beleid van het Rijk ('de rijksprestatie') geen effect heeft gehad bij lagere overheden. Bij onderzoek naar verklaringen voor het niet (voldoende) realiseren van prestaties en effecten zal vooral ook naar deze mogelijke verklaringsgrond moeten worden gekeken (zie ook hoofdstuk 6). Met het onderzoek wordt immers primair de minister aangesproken.

Meer in het algemeen geldt: wanneer de prestaties van decentrale actoren worden onderzocht, zou het onderzoek gericht moeten zijn op het vellen van een oordeel over de minister. Hiervoor is inzicht nodig in de relatieve bijdrage van het Rijk aan de

geleverde prestaties door decentrale actoren. Andere zaken waar de minister op aangesproken kan worden zijn:

- de kwaliteit van de informatie die hij/zij heeft over de realisatie van de gestelde doelen (inclusief de informatie die van de decentrale actoren komt of zou moeten komen);
- de mate waarin hij/zij de decentrale actoren, gegeven de doelen, aanspoort om bepaalde prestaties te leveren;
- de mate waarin hij/zij een systeem bouwt om te zorgen dat de prestaties en effecten geleverd kunnen worden (het scheppen van voorwaarden) en dat daarover informatie wordt verzameld.

7.7 Indicatoren voor resultaten van beleid

Er zijn verschillende manieren om resultaten van beleid (prestaties en effecten) te meten. De meest voor de hand liggende is natuurlijk geld, bijvoorbeeld kosten en opbrengsten. Een indicator die ook regelmatig wordt gebruikt, is de quality-adjusted life year (qaly). In de gezondheidszorg is dit een belangrijke indicator, bijvoorbeeld bij de besluitvorming over medische ingrepen. Maar ook op andere beleidsterreinen kan deze indicator worden toegepast, zoals bij milieuvraagstukken (welke winst in qalys levert de aanpak van luchtverontreiniging op?) of op justitiegebied (wat is de winst in qalys bij een vroegtijdige aanpak van huiselijk geweld?). Deze indicator is daarom heel geschikt om de maatschappelijke gevolgen van een maatregel te onderzoeken. Een andere indicator is gedrag. Veel beleidsmaatregelen zijn impliciet of expliciet gericht op verandering van gedrag. Dat kun je op verschillende manieren meten. Zie hiervoor ook hoofdstuk 9.

Een ander type indicator met een breder palet aan resultaten zijn indices, waarvan de brede welvaartsindex en de SDG-scores bekende voorbeelden zijn. In deze indices worden bijvoorbeeld niet alleen financiële resultaten, maar ook de resultaten voor milieu, welzijn, sociale percepties, ontplooiing en diversiteit meegenomen.

7.8 Onderzoekvarianten

7.8.1 Inleiding

We kunnen veel mogelijke onderzoeksvarianten onderscheiden, afhankelijk van de doelstelling, probleemstelling en onderzoeksvragen van je onderzoek. In dit hoofdstuk onderscheiden we drie algemene onderzoeksvarianten: doelrealisatie, doeltreffendheid en doelmatigheid.

De keuze van deze onderzoeksvariant(en) wordt in de eerste plaats bepaald door strategische overwegingen, zoals:

- mogelijkheid om gebruik te maken van de unieke bevoegdheden van de Algemene Rekenkamer;
- financieel belang van het te onderzoeken beleid;
- maatschappelijke relevantie en urgentie;
- potentiële risico's in de (uitvoerings)praktijk voor burgers en bedrijven;
- complexiteit van het veld of de lengte van de beleidsketen;
- mogelijkheid om bij te dragen aan verbetering;
- mogelijkheid om uitspraken te doen over doelmatigheid.

Daarnaast moet je rekening houden met een aantal uitvoeringstechnische factoren, zoals:

- politieke context;
- kwaliteit van de te onderzoeken beleidsdoelstelling;
- beschikbaarheid van onderzoeksmateriaal en beleidsinformatie;
- beschikbaarheid van onderzoekscapaciteit (kwantitatief en kwalitatief).

7.8.2 Wanneer beperk je je tot doelrealisatie?

Onderzoek naar de vraag of gestelde doelen gerealiseerd zijn, kan vooral belangrijk zijn als hierover onduidelijkheid of schijn duidelijkheid heerst en als dit verbonden is met discussies over uitgevoerd of uit te voeren beleid. Is de criminaliteit in een bepaalde periode toegenomen of afgenomen? Was de lucht anno 2019 schoner of vuiler dan in 2014? Zijn er in 2020 meer of minder jongeren zonder diploma van school gegaan dan in 2015?

Behalve op zo'n breed maatschappelijk niveau kan de vraag naar doelrealisatie ook belangrijk zijn als het 'alleen maar' gaat om geleverde diensten of producten (prestaties). De politieke context kan het zinvol maken om na te gaan hoeveel kilometer dijk er in een bepaalde periode is geïnspecteerd, hoeveel uren onderwijs er is gegeven, hoeveel uren de politie heeft gesurveilleerd of hoeveel WAO-uitkeringen zijn verstrekt. Meer dan eens blijkt het antwoord op deze ogenschijnlijk eenvoudige vragen helemaal niet makkelijk te vinden, terwijl beleidsmakers en/of volksvertegenwoordigers wel van bepaalde aannamen uitgaan.

Als de benodigde gegevens niet in een nette, bewerkte vorm in de beleidsinformatie zijn te vinden, is dat op zichzelf al een bevinding. Om verder te gaan, kun je proberen een samenhangend beeld uit verschillende deelbestanden, archieven en bestaande onderzoeken te construeren. Als ook dat niet kan en je wilt de conclusies desondanks niet beperken tot de gebrekkigheid van de informatie, zul je zelf gegevens moeten verzamelen. Voor dit type vragen heb je doorgaans relatief grootschalige informatie nodig, zeker als het om het niveau van de maatschappelijke veranderingen gaat.

Soms is het nog te vroeg of om andere redenen onmogelijk om vast te stellen of de einddoelen van beleid zijn bereikt. Dan kan het nuttig zijn om te onderzoeken in hoeverre tussendoelen zijn gerealiseerd. Soms is dat ook van belang omdat het realiseren van een einddoel niet onder de verantwoordelijkheid van een minister valt, maar het realiseren van bepaalde tussendoelen wel. Dit geldt in het bijzonder voor 'doorwerkingsonderzoek'. Het gaat hierbij om de centrale vraag of de voorwaarden hebben gewerkt die de minister gecreëerd heeft voor een goede uitvoering van het rijksbeleid op regionaal, provinciaal of gemeentelijk niveau.

Bij het onderzoek naar doelrealisatie stel je primair toetsende vragen, bijvoorbeeld: in hoeverre is het doel 'veilige dijken' gerealiseerd? De norm is dan dat de dijken veilig zijn. Om inzicht te krijgen in de feitelijke situatie, stel je beschrijvende vragen, bijvoorbeeld: hoeveel kilometer dijk is er in een bepaalde periode geïnspecteerd? Als blijkt dat de dijken niet veilig zijn of, wat vaker zal voorkomen, dat het niet duidelijk is of de dijken veilig zijn, dan kun je verklarende vragen stellen, bijvoorbeeld: wat is de oorzaak van de onduidelijkheid over de veiligheid van de dijken?

Het realiseren van een beleidsdoel is in veel gevallen afhankelijk van gedragsverandering bij een doelgroep van het beleid. Beleid kan in dergelijke gevallen alleen resultaten teweegbrengen wanneer de juiste personen op de een of andere manier iets merken van het beleid of ermee in contact komen.

Je kunt dan onderzoek doen naar *doelgroepbereik*. Dit is vooral belangrijk als er gereede twijfel is of belangrijke voorlichtingscampagnes of regelingen voldoende bekend zijn bij de doelgroep. Als dat wel zo is, weet je nog steeds niet of ze werken. Maar als het niet zo is, mag je het tegendeel gerust aannemen.

Mocht je, anders dan verwacht, vaststellen dat de doelgroep wel in voldoende mate wordt bereikt dan is eigenlijk een nieuwe afweging nodig: is dit voor ons voldoende of willen we verder kijken naar de effecten in termen van veranderingen in kennis, houding en gedrag bij de doelgroep?

Voor het pure doelgroepbereik zal je vaak met surveymethoden moeten werken. Als je verder wilt gaan en ook effecten wilt vaststellen, moet je doeltreffendheidsonderzoek doen. Dan is het nodig (maar ook lastig) om vergelijkbare gegevens te krijgen over de situatie voor en na (of met en zonder) de campagne.

Bij het onderzoek naar doelgroepbereik stel je primair toetsende vragen (zie hoofdstuk 5), bijvoorbeeld: in hoeverre is de doelgroep 'kleine ondernemers' bereikt met voorlichting over subsidiemogelijkheden? De norm is dan dat de doelgroep kennis heeft genomen van de subsidiemogelijkheden. Om inzicht te krijgen in de feitelijke situatie, stel je beschrijvende vragen (zie hoofdstuk 4), bijvoorbeeld: hoeveel kleine ondernemers hebben kennisgenomen van de subsidiemogelijkheden? Als blijkt dat 20% van de doelgroep is bereikt, dan kun je verklarende vragen stellen (zie hoofdstuk 6), bijvoorbeeld: wat is de oorzaak van het beperkte bereik van de doelgroep?

7.8.3 Wanneer doe je doeltreffendheidsonderzoek?

Doeltreffendheidsonderzoek is bij de Algemene Rekenkamer onderzoek naar de vraag in hoeverre dankzij het regeringsbeleid de gewenste beleidsresultaten worden bereikt. Onder beleidsresultaten worden zowel prestaties (output) als de daarmee gerealiseerde effecten (outcome) verstaan.

Doeltreffendheidsonderzoek dat verder gaat dan het bewerken van bestaande beleidsinformatie doet de Algemene Rekenkamer zelf als:

- de vraag niet op basis van bestaande beleidsinformatie is te beantwoorden;
- het niet te verwachten valt dat het desbetreffende ministerie of anderen hier op afzienbare termijn gedegen onderzoek naar (laten) doen;
- het van groot maatschappelijk en/of financieel belang is om te weten of het desbetreffende beleid tot resultaten leidt;
- de Algemene Rekenkamer in staat is om het benodigde onderzoek uit te voeren.

Dit alles moet je uiteraard nagaan als je het onderzoeksvoorstel schrijft.

In doeltreffendheidsonderzoek moet je altijd op twee vragen antwoord geven:

1. Zijn de met het beleid beoogde prestaties of effecten daadwerkelijk gerealiseerd? (Dit is dezelfde vraag die gesteld wordt in onderzoek naar doelrealisatie);
2. In hoeverre is het aan het beleid te danken dat de doelen bereikt zijn, of juist niet bereikt zijn?

Om op deze beide vragen te kunnen antwoorden, moet je aan informatie komen over de door het beleid te beïnvloeden situatie of gedragingen, liefst op verschillende tijdstippen. Denk bijvoorbeeld aan het volgen van scholing door jongeren voor en een tijdje na de introductie van stimulerend beleid. Ook moet je een solide beeld construeren van andere factoren die op deze zelfde situatie inwerken (bijvoorbeeld de situatie op de arbeidsmarkt) en een manier vinden om de invloed van het beleid te onderscheiden van de andere factoren. Vergelijkingen trekken tussen verschillende tijdstippen (voor en na nieuw beleid) en/of verschillende groepen (die meer en minder met het beleid in aanraking zijn gekomen) is bij de beantwoording van bovenstaande vragen essentieel.

In de praktijk is dit vaak niet haalbaar en zal je veelal moeten terugvallen op minder fraaie oplossingen, zoals een meting op één tijdstip en vergelijking met ander materiaal over de voorgaande situatie. Je loopt dan het risico dat vroeger gemaakte keuzen in vraagstelling en selectie van ondervraagden niet goed aansluiten bij jouw probleemstelling. Ook is het mogelijk om mensen te bevragen over de situatie voor de introductie van het nieuwe beleid. Dan heb je natuurlijk te maken met het risico dat het geheugen niet altijd betrouwbaar is en mensen sociaal wenselijke antwoorden geven.

Een standaardoplossing is er niet. Wel zijn er diverse manieren om de risico's in te perken. Die komen steeds neer op het motto: weet wat je vergelijkt. Let dus goed op

de samenstelling van de onderzochte groepen en zoek naar methoden om belangrijke verschillen te corrigeren.

Bij het onderzoek naar doeltreffendheid stel je primair toetsende vragen (zie hoofdstuk 5), bijvoorbeeld: in hoeverre heeft het beleid geleid tot de beoogde afname van het aantal werklozen met 30%? De norm is dan dat het aantal werklozen met 30% moet zijn afgenomen als gevolg van het gevoerde beleid. Om inzicht te krijgen in de feitelijke situatie, stel je beschrijvende vragen (zie hoofdstuk 04), bijvoorbeeld: wat is de ontwikkeling van het aantal werklozen in de periode dat het beleid van kracht is geweest? Als blijkt dat het aantal werklozen met 20% is afgenomen, kun je verklarende vragen stellen (zie hoofdstuk 6). Bijvoorbeeld: wat is de oorzaak van de beperktere afname van het aantal werklozen en welke invloed heeft het gevoerde beleid hierop gehad?

7.8.4 Wanneer doe je doelmatigheidsonderzoek?

Voor de burger telt niet alleen of beleidsresultaten zijn geleverd, maar ook tegen welke kosten dit is gebeurd en of dat beter kan. Daarom is het van belang om – in het verlengde van onderzoek naar effecten en prestaties – onderzoek te doen naar de doelmatigheid.

In onderzoek naar doelmatigheid vragen we ons af in hoeverre de beschikbare middelen (personeel, materieel, geld) optimaal worden benut. We onderscheiden drie soorten doelmatigheid (zie figuur 6.1), namelijk de relatie tussen middelen en:

- verrichte activiteiten om het beleid uit te voeren (doelmatigheid van de beleidsuitvoering);
- geleverde prestaties (doelmatigheid van de beleidsprestaties of efficiency);
- gerealiseerde effecten (doelmatigheid van de beleidseffecten of kosteneffectiviteit).

Het gaat daarbij om de vraag:

- of de uitvoering van het beleid, de geleverde prestaties en/of de gerealiseerde effecten met de inzet van minder middelen gerealiseerd hadden kunnen worden, dan wel
- of een betere uitvoering, méér prestaties en/of méér effecten verwezenlijkt hadden kunnen worden met dezelfde inzet van middelen.

Doelmatigheidsonderzoek gaat dus nog een stap verder dan doeltreffendheids-onderzoek. Je kijkt niet alleen naar de (kwaliteit van de) uitvoering, de geleverde prestaties of de gerealiseerde effecten, maar legt ook de relatie met de middelen die daarvoor zijn ingezet.

Doelmatigheid is op twee manieren een relatief begrip. Ten eerste gaat het om de onderlinge relatie tussen de middelen, de uitvoering, de prestaties of de effecten en de kwaliteit daarvan. Ten tweede kan doelmatigheid alleen worden bepaald in relatie

tot een referentiepunt, een organisatie is [on]doelmatiger dan een andere organisatie of dan in het verleden.

In bepaalde gevallen kun je wel een uitspraak over doelmatigheid doen zonder te vergelijken met anderen en/of het verleden. Als je namelijk aanzienlijke verspillingen aantreft (een gebrek aan zuinigheid) dan weet je dat de prestaties of effecten goedkoper, dus doelmatiger, geleverd hadden kunnen worden. Mogelijke vragen die je bij de keuze van een concreet onderzoeksonderwerp kunt stellen:

- Hoe belangrijk is een mogelijke doelmatigheidswinst?
- Is onvoldoende informatie over doelmatigheid beschikbaar?
- Gaat niemand anders er op korte termijn voor zorgen dat er voldoende informatie komt?
- Kan de rekenkamer daar wel voor zorgen?

Bij het onderzoek naar doelmatigheid stel je primair toetsende vragen (zie hoofdstuk 5). Een voorbeeld is: in hoeverre wordt het jaarlijks onderhoud van de YPR-pantserrijsvoertuigen doelmatig uitgevoerd? De norm is in dit geval dat het onderhoud is uitgevoerd tegen de laagst mogelijke kosten. Om inzicht te krijgen in de feitelijke situatie, stel je beschrijvende vragen (zie hoofdstuk 4), bijvoorbeeld: wat waren de kosten van onderhoud per pantservoertuig per jaar? Vervolgens zoek je naar vergelijkingen, bijvoorbeeld de kosten van onderhoud van vergelijkbare voertuigen of de kosten van onderhoud in andere landen. Als blijkt dat de kosten relatief hoog zijn, dan kun je verklarende vragen stellen (zie hoofdstuk 06). Een voorbeeld: wat is de oorzaak dat de kosten van het jaarlijks onderhoud relatief hoog zijn?

Een voorbeeld van een rekenkameronderzoek dat indicaties geeft voor de doelmatigheid van beleidseffecten is [Fiscale stimulering van elektrische auto's](#). We hebben onderzocht "of het bedrag van de fiscale stimulering niet te hoog is in relatie tot de vermindering van de CO₂-uitstoot, zowel per auto als in totaal." Ook in onze verantwoordingsonderzoeken over 2013 en 2014 is deze vraag gesteld. We hebben toen geconcludeerd dat de fiscale stimulering van zuinige auto's, gegeven de door het ministerie van Financiën berekende belastingderving van € 1.000 per vermeden ton CO₂, relatief duur is en dat het kabinet vermoedelijk meer CO₂ had kunnen besparen (dus meer effect had kunnen realiseren) bij een andere besteding van dat geld.

8.

Algemene normen voor DDO

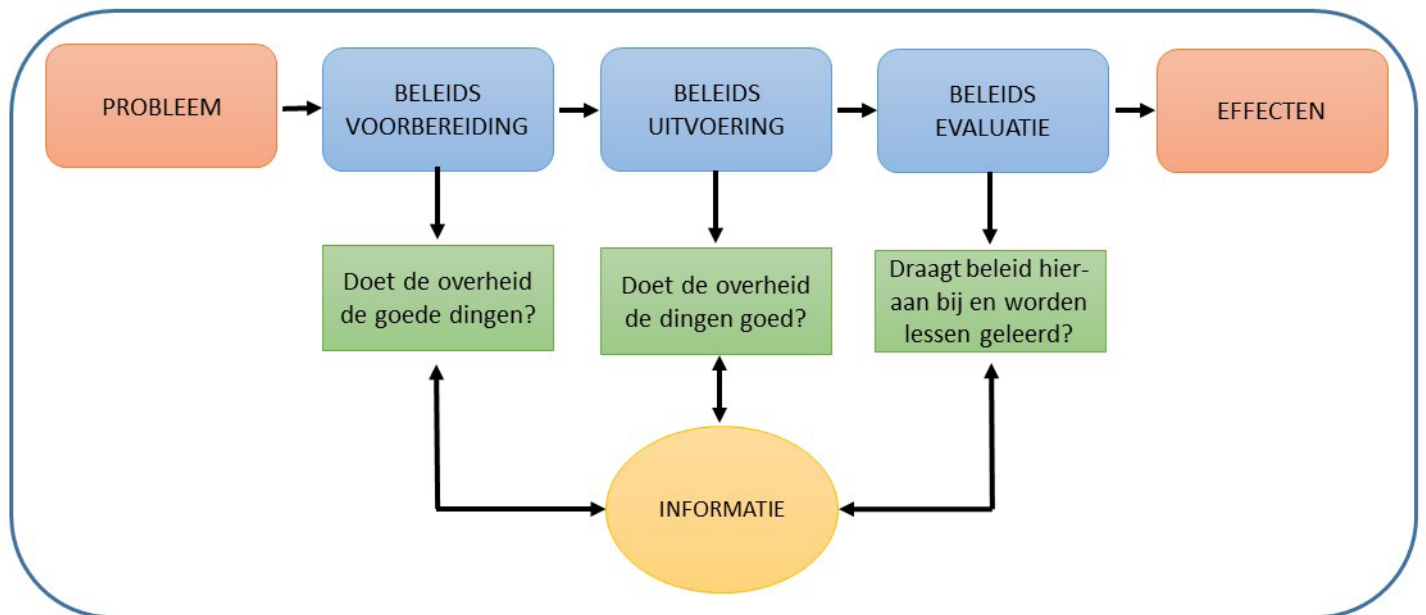
8.1 Inleiding

In dit hoofdstuk vind je de algemene normen die de Algemene Rekenkamer hanteert bij haar onderzoek. We gaan er daarbij van uit dat in de eerste plaats de ministers verantwoordelijk zijn voor de doelmatigheid en doeltreffendheid van hun beleid. Zij moeten bij voorstellen, voornemens en toezeggingen een toelichting geven waarin wordt ingegaan op de doelstellingen, de doeltreffendheid en de doelmatigheid die worden nagestreefd, de beleidsinstrumenten die worden ingezet en de financiële gevolgen voor het Rijk en, waar mogelijk, de financiële gevolgen voor maatschappelijke sectoren (CW 2016, art. 3.1). Ook moeten zij toezicht houden op de uitvoering van het beleid, en de doeltreffendheid en doelmatigheid van het beleid periodiek onderzoeken (CW 2016, art.4.1).

In [ISSAI 300](#) (Performance Audit Principles) wordt gesteld dat Auditors should establish suitable criteria which correspond to the audit questions and are related to the principles of economy, efficiency and effectiveness (ISSAI 300, § 27). Meer hulp bij het ontwikkelen van normen kun je vinden in [GUID 3910](#) (Central concepts for performance auditing, § 55 t/m 60) en [GUID 3920](#) (The performance auditing process, § 38 t/m 43).

De basisnormen voor het onderzoek naar doelmatigheid en doeltreffendheid van beleid kunnen worden gestructureerd aan de hand van de beleidscyclus: voorbereiding, uitvoering en evaluatie.

Figuur 8.1 Schematische weergave van de beleidscyclus: voorbereiding, uitvoering en evaluatie

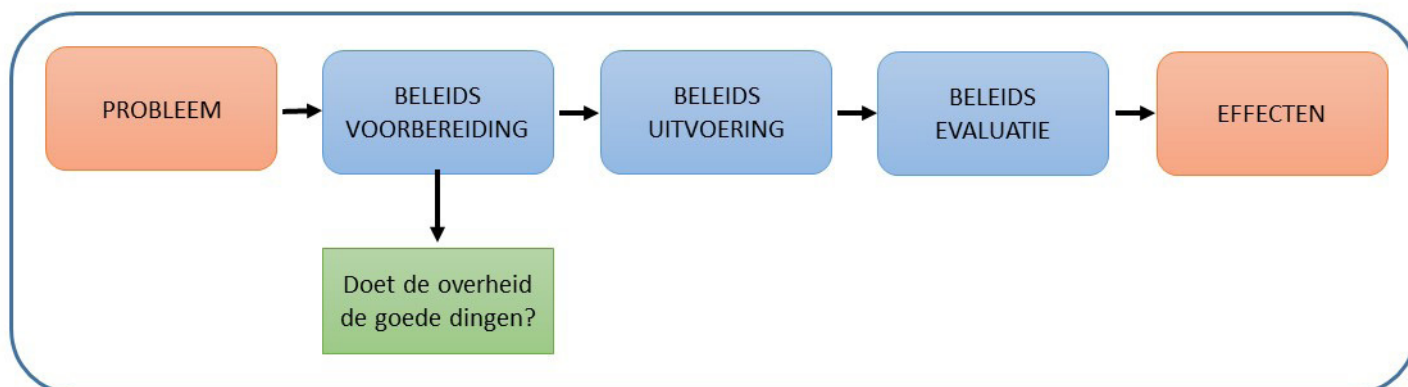


In dit hoofdstuk vind je basisnormen voor beleidsvoorbereiding, beleidsuitvoering, beleidsevaluatie, voor het gebruik van beleidsinformatie door gecontroleerden, voor kwaliteit van derdeninformatie voor gebruik in eigen onderzoek, voor beoordeling van de overheidsrol in beleidsnetwerken en voor toezicht op naleving en handhaving van EU-wet- en -regelgeving. In een bijlage wordt gedetailleerder ingegaan op de beoordeling van beleidsdoelen.

Let op: dit hoofdstuk is bedoeld als ondersteuning bij het toepassen van normen in je onderzoek. De meeste normen in dit hoofdstuk zijn afgeleid van eerdere onderzoeken van de Algemene Rekenkamer en het is de bedoeling dat je nagaat welke normen voor jouw specifieke onderzoek en jouw specifieke onderzoeksvragen relevant zijn. Dat is altijd maatwerk, zoals elk onderzoek maatwerk is. Tegelijkertijd zijn de normen niet vrijblijvend: het is belangrijk dat je de keuze van je normen verantwoordt in je projectvoorstel.

8.2 Basisnormen beleidsvoorbereiding

Figuur 8.2 *Beleidsvoorbereiding: doet de overheid de goede dingen?*



8.2.1 Uitgangspunten

De kwaliteit van de beleidsvoorbereiding is voor het Rijk het eerste en misschien wel het belangrijkste aangrijpingspunt voor beter beleid en betere uitvoering. Als het beleid niet goed wordt voorbereid, bestaat al voor de start van de uitvoering de kans op een slechte aansluiting tussen beleidsambities en praktijk. De resultaten van het beleid blijven dan waarschijnlijk achter bij wat in het vooruitzicht is gesteld.

In ons onderzoek zien we vaak dat een gebrekkige beleidsvoorbereiding, met name het ontbreken van een ex ante-evaluatie, een van de belangrijke verklaringen is voor achterblijvende beleidsresultaten. Zo ontbreken vaak heldere beleidsdoelen en is het lang niet altijd duidelijk of en hoe voorgestelde beleidsinstrumenten zullen bijdragen aan het oplossen van maatschappelijke problemen. Ook wordt de uitvoering niet altijd betrokken en niet gehoord bij het opstellen van beleid. Daardoor bestaat de kans dat uitvoerders worden opgezadeld met beleid dat niet werkt, slecht uitvoerbaar is en dat onevenredig grote administratieve lasten met zich meebrengt.

In de fase van beleidsvoorbereiding staat allereerst een drietal samenhangende uitgangspunten centraal: vraaggerichtheid, bewezen of potentiële werkzaamheid en duidelijke ambities. De vraag naar de uitvoerbaarheid en handhaafbaarheid van het beleid is vervolgens de belangrijkste schakel tussen beleid en uitvoering en het sluitstuk van de beleidsvoorbereiding. Het gaat hierbij om de vraag of de gekozen maatregelen ook praktisch uitgevoerd en gehandhaafd kunnen worden: gaat de beleidsmaker na of de daartoe benodigde capaciteiten en kwaliteiten daadwerkelijk aanwezig zijn bij de uitvoerder van het beleid? Is het realistisch dat de beleidsdoelen zullen worden gehaald binnen de gestelde randvoorwaarden qua tijd en geld? En niet onbelangrijk: gaat de beleidsmaker zelf na of het beleid uitvoerbaar en handhaafbaar is?

Als in de ontwerpfase van het beleid onvoldoende aandacht is voor de benodigde inzet van tijd, geld, mensen en middelen, dan is de kans groot dat het beleid niet tot de gewenste resultaten zal leiden. De uitvoering zal dan doen wat kan en niet wat moet.

8.2.2 Normen

Hoofdnorm: Beleid is zorgvuldig voorbereid, uitvoerbaar en handhaafbaar.

8.2.2.1 Zorgvuldig

- *Vraaggericht beleid*: de overheid baseert haar beleid op vragen uit de samenleving. Het beleid sluit aan op de behoeften en problemen.
- *Potentieel of bewezen werkzaam*: de overheid moet zich afvragen met welke (potentieel werkzame of bewezen werkzame) beleidsinstrumenten de beleidsdoelstellingen het best kunnen worden bereikt. Het beleid moet daarom gebaseerd zijn op een geëxpliciteerde beleidstheorie, het geheel aan veronderstellingen dat aan een beleid ten grondslag ligt.

Let op: Het beoordelen (of reconstrueren) van een beleidstheorie is een hulpmiddel om je onderzoek uit te voeren. Het kan je inzichten bieden in mogelijke oorzaken voor achterblijvende beleidsresultaten. Het is nadrukkelijk geen blauwdruk of normatief uitgangspunt voor het beleid zelf. Net zomin is deze handleiding een pleidooi voor een uitsluitend rationalistische kijk op beleid.

Voor het beoordelen van de kwaliteit van de beleidstheorie kun je de volgende criteria gebruiken:

- De beleidstheorie moet expliciet beschreven zijn, hetzij in beleidsdocumenten, hetzij door een verwijzing naar relevante (wetenschappelijke) literatuur. De beleidstheorie moet de doelstellingen van het desbetreffende beleid bevatten.
- De beleidstheorie moet valide en plausibel zijn: de veronderstelde gedragsmechanismen bij de doelgroep van het beleid moeten overeenkomen met de huidige kennis over deze mechanismen binnen de sociale, gedrags- en economische wetenschappen. Ook de causale mechanismen in de beleidstheorie moeten plausibel zijn, bijvoorbeeld door aannemelijk te maken dat het beleid bijdraagt aan de oplossing van de gesignaleerde (maatschappelijke) problemen. Het begrip plausibiliteit is iets ruimer dan validiteit: het omvat behalve onderbouwing door empirische gegevens ook logica en andere vormen van wetenschappelijke onderbouwing.
- De veronderstellingen waaruit de theorie is opgebouwd, mogen niet tegenstrijdig zijn en moeten een logische samenhang vertonen.
- De beleidstheorie moet gericht zijn op variabelen die door het beleid kunnen worden beïnvloed.

- De gebruikte begrippen moeten helder en ondubbelzinnig omschreven zijn.
- Alle te verwachten beleidseffecten moeten in ogenschouw zijn genomen, inclusief mogelijke neveneffecten of averechtse effecten.
- Het moet helder zijn wat van alle bij de uitvoering van het beleid betrokken actoren wordt verwacht. Deze verwachtingen moeten realistisch zijn, dat wil zeggen dat er voldoende tijd en capaciteit (mensen, middelen en geld) beschikbaar moeten zijn.
- Het beleid moet nog zinvol zijn tegen de tijd dat het effect sorteert.
- De overheid moet duidelijke ambities hebben. Dat is nodig om vast te stellen wat nodig is (in termen van tijd, mensen, middelen en geld) om de beleidsdoelen te realiseren. Daarom moeten beleidsdoelen SMART-C geformuleerd zijn: specifiek, meetbaar, afgestemd, realistisch, tijdgebonden en consistent.

Norm	Uitwerking
Specifiek	Bij beleidsdoelen moet duidelijk zijn omschreven welke prestaties en effecten worden beoogd en welke doelgroepen (zo die er zijn) moeten worden bereikt
Meetbaar	Gemeten moet kunnen worden of beleidsdoelen al dan niet bereikt zijn. Daarvoor zijn eenduidige begrippen vereist. Vaak is ook een nulmeting nodig.
Afgestemd	Beleidsdoelen moeten zijn afgestemd met de Tweede Kamer en met de uitvoerders van het beleid. Ook tussentijdse bijstellingen of wijzigingen van de doelstelling moeten met hen worden afgestemd.
Realistisch	Beleidsdoelen moeten zodanig zijn gekozen dat ze realiseerbaar zijn onder alle plausibele omstandigheden (exogene variabelen en beleidsmaatregelen).
Tijdgebonden	In beleidsdoelen moet zijn vastgelegd wanneer het einddoel moet zijn bereikt. Ligger de doelen ver weg, zoals bij langlopende projecten, dan moet een tijdspad met gespecificeerde mijlpalen zijn aangegeven (bijvoorbeeld tussendoelen en bijbehorende data, of het tempo waarin doelen moeten zijn gerealiseerd).
Consistent	Doelen mogen niet strijdig zijn met de doelen van ander beleid. Tegenstrijdige doelen frustreren de effectiviteit van het overheidsbeleid. Doelen moeten aansluiten op de basisgegevens waarop ze gebaseerd zijn.

In de bijlage worden de normen SMART-C uitvoeriger omschreven.

Let op: Als de overheid wil bijdragen aan het oplossen van problemen, dan is het niet altijd mogelijk op voorhand SMART-C-doelstellingen te formuleren. Ook is het passende beleidsinstrument niet altijd onmiddellijk voorhanden. Het is van belang om dit zowel in het beleid, als in het onderzoek van beleid te onderkennen. Als er geen SMART-C-doelstellingen geformuleerd zijn, kun je nagaan of dat plausibel is.

- Als het niet plausibel is dat SMART-C-doelen **ontbreken**, dan kun je je onderzoek daarop richten en moet je goed onderbouwen waarom de overheid wel degelijk SMART-doelstellingen had kunnen formuleren. Eventueel kun je, als onderdeel van het onderzoek, proberen alsnog doelen te reconstrueren in overleg met gecontroleerden en/of stakeholders.

- Als het ontbreken van SMART-C-doelen **wel** plausibel is, kun je nagaan of de overheid op een andere wijze haar ambities heeft geformuleerd. Zo wordt beleid steeds vaker voorbereid en uitgevoerd in samenspraak met andere partners (gemeenten, provincies, openbare organen, maatschappelijke organisaties, bedrijven) die hun eigen belangen en doelen hebben. In die gevallen kun je niet verwachten dat de overheid doelen SMART-C formuleert, maar richt je je onderzoek meer op de eigen rol van de overheid in dergelijke netwerken. Zie hiervoor verder paragraaf 8.7.

Wijziging van beleidsdoelen

Beleidsdoelen kunnen om verschillende redenen in de loop van het proces veranderen. Wanneer dit gebeurt, moet dit goed worden onderbouwd en worden verantwoord aan de Tweede Kamer. De geherformuleerde doelen moeten, voor zover mogelijk en realistisch, opnieuw aan bovengenoemde normen voldoen.

8.2.2.2 Uitvoerbaar

De ambities van het beleid, de tijd, de mensen en de middelen en het geld zijn met elkaar in balans. Met andere woorden: er moet sprake zijn van een kloppende 'vierkantsvergelijking'. De doelen kunnen gerealiseerd worden binnen de afgesproken tijd met de beschikbare mensen en middelen. We verwachten dat de overheid en daarmee verbonden organisaties bij de totstandkoming van beleid nagaan of het beleid (ook in tijd) uitvoerbaar is.

In dit verband is ook de kwaliteit van ramingen belangrijk. De raming moet op betrouwbare wijze tot stand zijn gekomen en in de beleidsvoorbereiding correct zijn gebruikt.

Normen voor de totstandkoming:

- Zijn de aannames waarmee de raming is opgesteld plausibel, objectief en consistent (dus niet sturend naar een gewenste uitkomst)?
- Zijn de gegevens (brondata) waarmee de raming is opgesteld relevant (dus niet sturend naar een gewenste uitkomst), juist en zo actueel mogelijk?
- Zijn eventueel, vooral naarmate de raming met meer onzekerheden is omgeven, scenario's of gevoeligheidsanalyses toegepast?

Normen voor het gebruik:

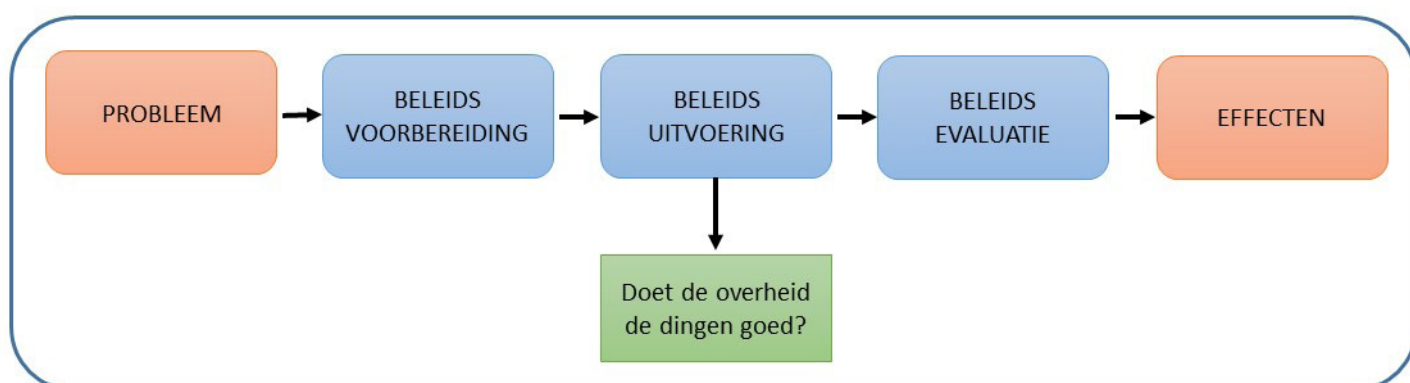
- Sluit de uiteindelijke beleidskeuze aan bij de raming?
- Hebben de gebruikers van de raming voldoende rekening gehouden met de onzekerheidsmarges die de raming omgeven?
- Als uiteindelijk een puntschatting gekozen is, is die keuze dan duidelijk onderbouwd en plausibel?

8.2.2.3 Handhaafbaar

De overheid moet vaststellen dat de handhavende instanties de nieuwe regelgeving kunnen handhaven. Hierbij wordt onder meer gekeken of de verwachte extra handhavingsinspanning toereikend is als de regelgeving niet spontaan zal worden nageleefd. Om te kunnen vaststellen of het beleid handhaafbaar is, worden inspecties en andere relevante handhavers bij de beleidsvoorbereiding betrokken.

8.3 Basisnormen beleidsuitvoering

Figuur 8.3 Beleidsuitvoering: dot de overheid de dingen goed?



8.3.1 Uitgangspunten

In deze fase moeten de uitvoerders van het beleid, gegeven de randvoorwaarden tijd, geld, mensen en middelen, bepaalde doelen zien te realiseren. De uitvoering moet op een effectieve en efficiënte wijze plaatsvinden. Denk hierbij bijvoorbeeld aan efficiënte werkprocessen, goede samenwerking, risicomangement, ICT-toepassingen, goed opgeleide mensen, acceptabele arbeidsproductiviteit, efficiënt gebruik van middelen. Efficiënt gebruik van middelen kan er zelfs voor zorgen dat eerder onhaalbare doelen (als gevolg van onvoldoende beschikbare tijd, geld, mensen en middelen) toch gerealiseerd kunnen worden.

De beleidsuitvoering moet ertoe leiden dat de doelen van het beleid worden gerealiseerd en dat dit op effectieve en efficiënte wijze gebeurt. Het gaat hierbij om de volgende aspecten:

- Bij *doelrealisatie* gaat het om de vraag of de doelstellingen van het beleid zijn gerealiseerd. Daarbij wordt geen causaal verband gelegd tussen de prestaties van de overheid en de mate waarin doelstellingen worden behaald. Op basis hiervan kunnen dus geen uitspraken worden gedaan over de mate waarin het beleid wel of niet gewerkt heeft.

- Die uitspraken zijn wel mogelijk bij onderzoek naar de *doeltreffendheid* van beleid: daarbij wordt nagegaan of doelstellingen zijn gerealiseerd dankzij de beleidsprestaties. Doordat er sprake kan zijn van een lange beleidsketen en doordat ook externe factoren van invloed kunnen zijn op de te bereiken doelen, is het vaak moeilijk om het directe causale verband tussen prestaties en effecten aan te tonen. Dat neemt niet weg dat je in elk geval kunt proberen in je onderzoek indicaties te vinden voor de doeltreffendheid van het beleid.

Bij *doelmatigheid* vragen we ons af in hoeverre de beschikbare middelen (personeel, materieel, geld) optimaal worden benut. We onderscheiden drie soorten doelmatigheid (zie figuur 7.1, namelijk de relatie tussen middelen en:

- verrichte activiteiten om het beleid uit te voeren (doelmatigheid van de beleidsuitvoering);
- geleverde prestaties (doelmatigheid van de beleidsprestaties of efficiency);
- gerealiseerde effecten (doelmatigheid van de beleidseffecten of kosteneffectiviteit).

Het gaat daarbij om de vraag:

- of de uitvoering van het beleid, de geleverde prestaties en/of de gerealiseerde effecten met de inzet van minder middelen gerealiseerd hadden kunnen worden, dan wel
- of een betere uitvoering, meer prestaties en/of meer effecten verwezenlijkt hadden kunnen worden met dezelfde inzet van middelen.

8.3.2 Normen

Hoofdnorm: Beleid wordt effectief en efficiënt uitgevoerd en is doeltreffend en doelmatig.

8.3.2.1 Uitvoering

- De beoogde prestaties worden geleverd en zijn toereikend voor het bereiken van de doelen.
- De beoogde prestaties worden tijdig geleverd.
- Van het budget en de middelen wordt efficiënt (zuinig) gebruik gemaakt.
- De geleverde prestaties dragen daadwerkelijk bij aan het bereiken van de doelen en de kosten staan in reële verhouding tot de baten. Dit betekent in elk geval dat de kosten de baten niet mogen overschrijden.
- Voor zover zij daar bevoegdheden heeft zorgt de overheid ervoor dat ook de doelen van het desbetreffende EU-beleid worden gerealiseerd.
 - Er wordt beleidsinformatie verzameld om vast te stellen of de uitvoering loopt zoals verwacht, of de resultaten worden bereikt en of het beleid doelmatig en doeltreffend wordt uitgevoerd.

- Op basis van deze informatie worden zo nodig aanpassingen doorgevoerd in het beleid en/of de uitvoering.
- De informatie is deugdelijk: juist, volledig, relevant, betrouwbaar, begrijpelijk en vergelijkbaar.
- Wanneer de overheid met andere actoren in beleidsnetwerken opereert draagt zij, in aanvulling op het bovenstaande, zorg voor een zodanige vervulling van haar rollen als medespeler of regisseur dat de benodigde beleidsinformatie wordt verzameld.

8.3.2.2 Doelrealisatie

- De beleidsdoelen moeten worden gerealiseerd. Als beleidsdoelen niet gerealiseerd dreigen te worden, heroverweegt de beleidsmaker de ambities of de inzet van mensen, middelen, geld of tijd.

Bij het beoordelen van de doelrealisatie gaat het om de beoogde prestaties en effecten, die worden afgezet tegen de gerealiseerde prestaties en effecten. De beoogde prestaties en effecten zijn dus de 'normen' waaraan wordt getoetst. De beoordeling hiervan moet wel redelijk zijn: als door onvoorziene omstandigheden de doelen niet worden gehaald, moeten we dit meewegen in ons oordeel.

Het kan voorkomen dat het bereiken van de einddoelen niet goed kan worden onderzocht. Je kunt dan tussendoelen onderzoeken. De mate van doelgroepbereik is een voorbeeld van zo'n tussendoel. Je bepaalt dan of de beoogde doelgroep wat van het beleid heeft gemerkt, maar nog niet of de beoogde doelen zijn gerealiseerd. Voor het doelgroepbereik kun je de volgende normen hanteren:

- De doelgroep van het beleid moet zijn omschreven en zo duidelijk mogelijk afgebakend. Ook moet de keuze van de doelgroep zijn gemotiveerd.
- De beoogde doelgroep van het beleid moet worden bereikt. Daartoe moet
 - de beoogde doelgroep worden afgezet tegen de doelgroep die na/bij uitvoering van het beleid is bereikt;
 - de beoogde 'vorm van bereik' van de doelgroep worden afgezet tegen de 'vorm van bereik' die is gerealiseerd. Een voorbeeld van een beoogde vorm van bereik is 'in contact komen met';
 - het beoogde tijdstip waarop de beoogde doelgroep in de beoogde vorm moet zijn bereikt, worden afgezet tegen het tijdstip waarop dit is gerealiseerd.

8.3.2.3 Doeltreffendheid

- Overheidsbeleid moet doeltreffend (effectief) zijn. Het gevoerde beleid leidt ertoe dat de beoogde effecten (zoals verwoord in de doelstellingen van het beleid) worden gerealiseerd.

- Er mag geen sprake zijn van (substantiële) negatieve neveneffecten. Zo moeten de volgende ongewenste effecten worden voorkomen:
 - Cadeaueffect: dit effect treedt onder andere op bij subsidies: de doelgroep zou het door de subsidie gestimuleerde gedrag ook zonder of met minder subsidie hebben vertoond.
 - Verdringingseffect: dit is een effect dat bijvoorbeeld optreedt bij instrumenten op het gebied van de arbeidsvoorziening: sommigen worden weliswaar aan werk geholpen, maar dit gaat ten koste van andere werknemers of werkzoekenden.
 - ‘Benefit snatching’: dit is het verschijnsel dat aanbieders van een te subsidiëren product (of anderen in de productkolom) profiteren van het geboden subsidievoordeel door de prijs te verhogen.

Zie ook hoofdstuk 7.

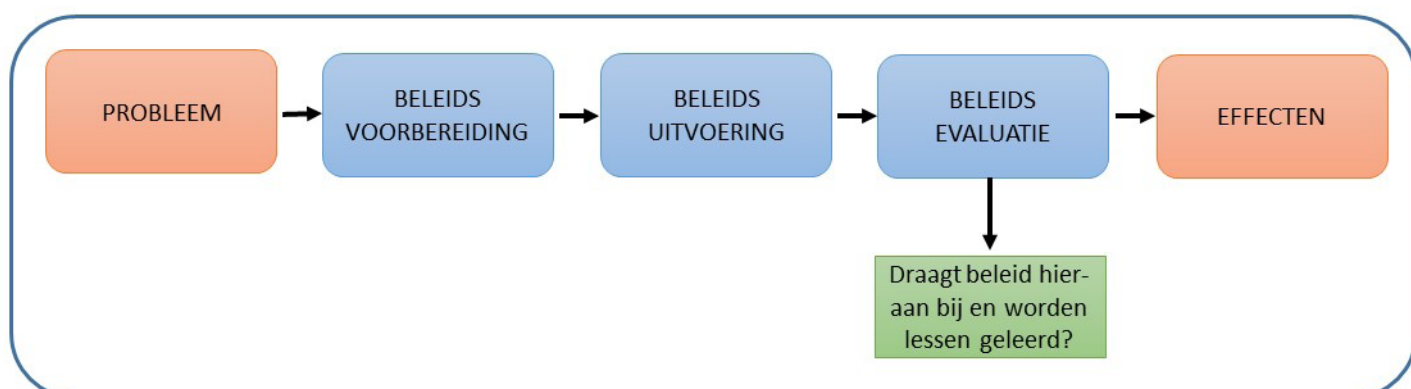
8.3.2.4 Doelmatigheid

- Overheidsbeleid dient uiteindelijk doelmatig te zijn.

Let op: doelmatigheid is een relatief begrip. De uitvoering, de prestaties en de effecten van het beleid van de minister zijn meer of minder doelmatig dan van een vergelijkbare organisatie en/of in een eerdere situatie. Dit betekent dat we nooit een absolute uitspraak kunnen doen over de doelmatigheid. Doelmatigheid is ook een dynamisch begrip: nieuwe inzichten en nieuwe technieken kunnen de doelmatigheid bevorderen. Zie ook hoofdstuk 7.

8.4 Basisnormen beleidsevaluatie

Figuur 8.4 Beleidsevaluatie: draagt beleid hieraan bij en worden lessen geleerd?



8.4.1 Uitgangspunten

In de fase van beleidsevaluatie moet de overheid nagaan of na laten gaan of de goede dingen zijn gedaan en of de dingen goed zijn gedaan. Het is daarom belangrijk dat er bruikbare beleidsevaluaties beschikbaar zijn die bijdragen aan het lerend vermogen van de overheid. Daarnaast verwachten we dat de overheid op deugdelijke en transparante wijze publieke verantwoording aflegt over de besteding van publiek geld.

8.4.2 Normen

Hoofdnorm: De doeltreffendheid en doelmatigheid van het beleid worden periodiek op transparante wijze geëvalueerd met het oog op versterking van het lerend vermogen.

8.4.2.1 Periodieke evaluatie

- Het ministerie beschikt over een werkend evaluatiesysteem dat voorziet in de evalueerbaarheid van het beleid, de monitoring van en het toezicht op de voortgang, de (kwaliteit van de) evaluatie en het benutten van de (gevalideerde) resultaten van de evaluatie.
- Alle beleidsterreinen van een ministerie die vallen onder de ministeriële verantwoordelijkheid, worden periodiek (in de regel ten minste eens in de 7 jaar) en systematisch geëvalueerd. Van deze bepaling kan worden afgeweken als het niet mogelijk is om zinvolle tussentijdse doelstellingen te formuleren (bij langlopend beleid) of als de kosten van evaluatieonderzoek te groot zijn in verhouding tot de baten ervan.
- Het evaluatieonderzoek naar de doeltreffendheid en doelmatigheid van het beleid voldoet aan de eisen van relevantie, validiteit en betrouwbaarheid. Zie hiervoor paragraaf 8.5.

8.4.2.2 Verantwoording

Het openbaar bestuur moet transparant functioneren. Dat betekent dat relevante informatie voor belanghebbenden toegankelijk moet zijn en dat begrotingen en jaarverslagen deugdelijk moeten zijn. De overheid moet, behalve over de financiën, ook publiek verantwoording afleggen over de effectiviteit en efficiëntie van het beleid en over de mate waarin publieke taken op een effectieve, efficiënte, responsieve en maatschappelijk verantwoorde wijze zijn uitgevoerd. De overheid moet eveneens verantwoording afleggen over het lerend vermogen van de organisatie en over de verbetermaatregelen die in gang zijn gezet. In het rapport [Verantwoorden over beleid](#) vind je specifieke normen voor (de kwaliteit van) verantwoordingsinformatie. De overheid (het kabinet) legt primair verantwoording af aan het parlement. We hanteren daarbij de norm dat het parlement tijdig, volledig en juist wordt

geïnfomeerd. Let op: dit is geen specifieke DDO-(basis)norm, maar een staatsrechtelijke norm die voortvloeit uit artikel 68 van de Grondwet.

8.4.2.3 Een lerende organisatie

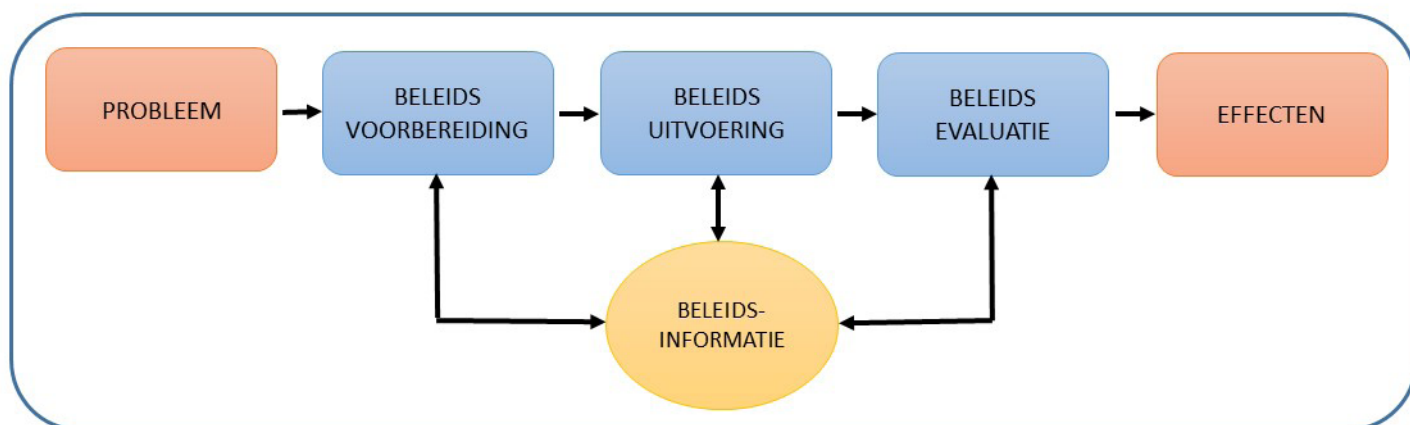
Bij het leren van evalueren gaat het om de beleidsveronderstellingen, de beleidsvoorbereiding en de uitvoering van het beleid. Hierbij komen idealiter alle relevante aspecten (tijd, mensen, middelen en geld) aan bod. Ook als het gaat om het monitoren en evalueren van beleid moet de overheid responsief zijn. Ze moet lerend vermogen tonen, bijvoorbeeld door prestatie- en effectindicatoren regelmatig te valideren. Het gaat bij het beoordelen van lerend vermogen om vragen als:

- Wordt het beleid bijgesteld als blijkt dat het niet of minder succesvol is?
- Wordt bij nieuw beleid geleerd van ervaringen uit het verleden?
- Heeft de overheid haar eigen lerend vermogen georganiseerd?

Een voorbeeld van een onderzoek naar lerend vermogen is [Inzicht in kwaliteit](#). Een ander voorbeeld is het onderzoek [Leren van subsidie-evaluaties](#).

8.5 Basisnormen gebruik beleidsinformatie door gecontroleerden

Figuur 8.5 Beleidsinformatie is van belang voor beleidsvoorbereiding, beleidsuitvoering en beleidsevaluaties



8.5.1 Uitgangspunten

Onder beleidsinformatie verstaan we gestructureerde en toegelichte informatie over:

- de doelstelling(en) van beleid;
- de beoogde en/of gerealiseerde effecten van beleid;
- de daartoe te leveren en/of geleverde prestaties;
- de daarmee gemoeide kosten.

Beleidsinformatie is dus gerelateerd aan alle fasen van het beleidsproces. Bruikbare beleidsinformatie is van groot belang voor het lerend vermogen van de overheid. Ze is ook nodig om het beleid goed te kunnen voorbereiden, uit te voeren en te evalueren. Bruikbare beleidsinformatie ondersteunt de bedrijfsprocessen en waarborgt democratische, juridische en historische waarden. Een goed functionerende informatiehuishouding – het totaal aan regels en voorzieningen gericht op de informatiestromen en op de opslag en/of archivering van informatie – is hiervoor een belangrijke voorwaarde.

8.5.2 Normen

Hoofdnorm: de door de gecontroleerde gebruikte beleidsinformatie moet deugdelijk tot stand komen en op deugdelijke wijze worden gepresenteerd en onderhouden.

Voor de beoordeling van de kwaliteit van de beleidsinformatie onderscheiden we de totstandkoming en de presentatie van informatie.

8.5.2.1 Totstandkoming van beleidsinformatie

- Beleidsinformatie moet betrouwbaar tot stand zijn gekomen.
- Begrippen moeten valide gemeten worden, dat wil zeggen: begrippen moeten zodanig geoperationaliseerd en gemeten zijn, dat je meet wat je wilt meten.
- Beleidsinformatie moet ordelijk en controleerbaar tot stand gekomen zijn. Het proces van totstandkoming moet achteraf te reconstrueren zijn.
- Het proces van totstandkoming van beleidsinformatie moet kosteneffectief zijn. De kosten van het genereren van de beleidsinformatie moeten opwegen tegen de opbrengsten ervan.

8.5.2.2 Presentatie van beleidsinformatie

- Relevantie:
 - De beleidsinformatie moet volledig zijn. Dat wil zeggen dat, gelet op de doelen van het beleid, alle informatie om te sturen, leren of verantwoorden beschikbaar moet zijn.
 - De beleidsinformatie moet actueel zijn.
 - De benodigde beleidsinformatie moet op tijd beschikbaar zijn.
- Transparantie:
 - De beleidsinformatie moet toegankelijk zijn voor de gebruikers ervan.
 - De beleidsinformatie mag niet voor meer dan één uitleg vatbaar zijn.
 - De beleidsinformatie moet evenwichtig (eerlijk, objectief en gebalanceerd) worden gebruikt. Is alle beschikbare relevante informatie gebruikt en op de juiste wijze gewogen? Is informatie niet gekleurd op grond van wat de minister zelf vond? Worden in de beleidsreactie van een minister bij een

evaluatie-rapport ook de minder positieve conclusies benoemd? Een voorbeeld van een beoordeling van het gebruik van beleidsinformatie is het onderzoek [Leren van subsidie-evaluaties](#).

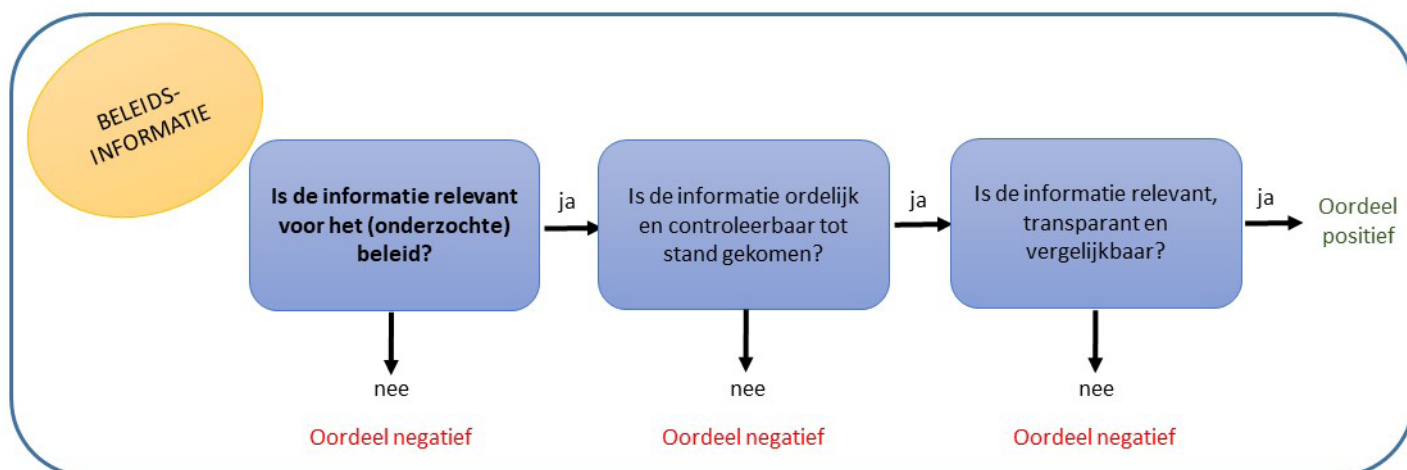
- De overheid moet inzicht geven in de kwaliteit van de gebruikte gegevens en de herkomst ervan. Wordt helder aangegeven hoe de informatie is gebruikt en welke beperkingen er eventueel aan kleven?
- **Vergelijkbaarheid:**
 - Consistentie in de tijd: de onderliggende gegevens van de beleidsinformatie moeten in de tijd vergelijkbaar zijn en wijzigingen in eerder verstrekte gegevens moeten worden geëxpliciteerd.
 - Consistentie tussen onderdelen van het beleidsproces: de gegevens moeten aansluiten bij de wijze waarop het beleid is gedefinieerd; de wijze van prestatiemeting moet aansluiten op de ingezette instrumenten en de prestatiedoelstelling; de wijze van effectmeting moet aansluiten bij het maatschappelijk probleem en de effectdoelstelling.
 - Consistentie tussen verschillende informatiebronnen: verschillende gegevens over eenzelfde object moeten onderling vergelijkbaar zijn. Als dit niet zo is, moet dat expliciet worden vermeld en zo mogelijk verklaard.

8.5.2.3 Informatiehuishouding

- De informatiehuishouding moet zo worden ingericht dat (digitale) informatie van meet af aan, maar ook na verloop van tijd raadpleegbaar, toegankelijk en authentiek is en blijft.

Bij het beoordelen van de kwaliteit van door de overheid gebruikte beleidsinformatie kun je het volgende schema gebruiken.

Figuur 8.6 Toetsen van beleidsinformatie aan de bijbehorende normen



Als het oordeel over de informatie negatief is, kun je de conclusie trekken dat gebruikgemaakt is van niet betrouwbare informatie voor de onderbouwing van beleid of voor verantwoording aan de Tweede Kamer.

In 08.06 beschrijven we de normen voor de kwaliteit van derdeninformatie voor gebruik in eigen onderzoek.

8.6 Normen voor de kwaliteit van derdeninformatie voor gebruik in eigen onderzoek

8.6.1 Inleiding

Veel van de informatie die wij in ons onderzoek gebruiken is afkomstig van andere instellingen, zoals uitvoeringsinstellingen, departementen, onderzoeksbureaus en planbureaus. Informatie die niet uit eigen onderzoek afkomstig is, noemen wij derdeninformatie. Onder informatie verstaan we hier zowel schriftelijke als digitale informatie. Dit kan heel uiteenlopende informatie betreffen zoals gegevens en/of conclusies uit onderzoeksrapporten, gegevensbestanden, verslagen, notulen en plattegronden. Het is uiteraard van belang dat deze informatie, net als zelf verzamelde informatie, van goede kwaliteit is. Ten eerste omdat we alleen kwalitatief goede informatie van andere partijen kunnen gebruiken als onderbouwing van conclusies in ons eigen onderzoek. In de tweede plaats omdat we zelf ook een oordeel vellen over de kwaliteit en het gebruik van door de overheid gepresenteerde informatie. Een voorbeeld hiervan is de niet-financiële informatie in de jaarverslagen of het onderzoek [Leren van subsidie-evaluaties](#).

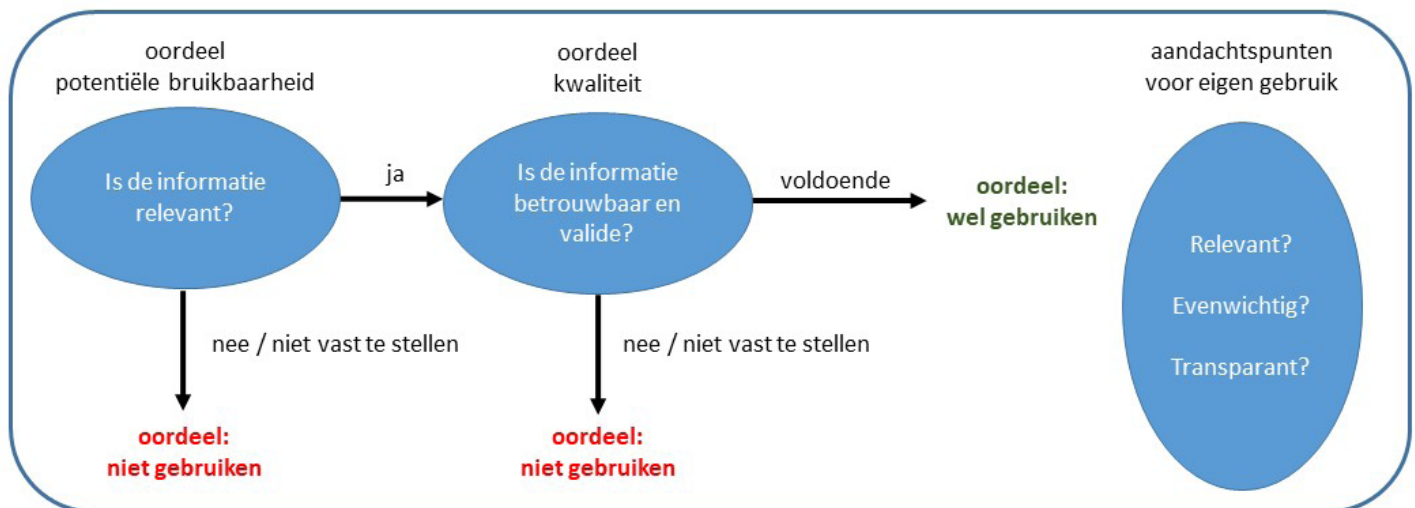
Hier gaan we in op de normen voor de bruikbaarheid en kwaliteit van informatie van derden voor gebruik in een rekenkameronderzoek.

De ISSAI's benadrukken dat aan de gebruiker van het rapport zekerheid moet worden gegeven dat onze bevindingen en conclusies gebaseerd zijn op betrouwbare en relevante informatie. Dat betekent dat we in onze publicatie (of in een bijlage bij die publicatie op onze website) transparant moeten zijn over hoe we tot die bevindingen en conclusies zijn gekomen. Dat betekent ook dat we duidelijk maken dat we alle in het onderzoek gebruikte derdeninformatie hebben beoordeeld op bruikbaarheid en kwaliteit. Als je informatie niet hebt beoordeeld of niet hebt kunnen beoordelen, dan gebruik je de informatie niet. Als er beperkingen zijn in de door ons gebruikte informatie, maken we dat ook duidelijk, inclusief de eventuele gevolgen voor onze bevindingen en conclusies. Bedenk daarbij overigens dat absolute zekerheid geven vrijwel nooit mogelijk is. Zie hiervoor ook [ISSAI 100](#), § 31, 49 en [ISSAI 300](#), § 21-23.

8.6.2 Stapsgewijs beoordelen van de bruikbaarheid en kwaliteit van derdeninformatie

Aansluitend bij de ISSAI's is de centrale vraag: kan ik deze gegevens gebruiken voor mijn onderzoek? Om praktisch te werk te gaan zou je de beoordeling stapsgewijs kunnen aanpakken. Deze aanpak wordt hieronder schematisch weergegeven. Hierbij is het van belang om in het achterhoofd te houden dat het oordeel na iedere stap gebaseerd is op je eigen expertise en ervaringen. Het is dus in zekere mate subjectief. Om dit enigszins te compenseren kun je een oordeel voorleggen aan een collega, of de beoordeling samen met een teamgenoot doen. Zeker in geval van twijfel is dit aan te raden.

Figuur 8.7 Stapsgewijs beoordelen van de bruikbaarheid en kwaliteit van derdeninformatie



Stap 1: Beoordelen (potentiële) bruikbaarheid: relevantie

Stel jezelf de vraag of de informatie die je wilt gaan gebruiken ook echt relevant is voor je onderzoek. Omdat je in de beginfase van je onderzoek vaak nog niet helemaal weet welke uitspraken je gaat doen of welke conclusies je eventueel gaat trekken, gaat het hier in eerste instantie om de potentiële relevantie. Bij het eindproduct zal de relevantie waarschijnlijk nogmaals en nu iets specifieker onder de loep moeten worden genomen: is deze informatie relevant voor het trekken van deze conclusie? Je kunt jezelf in het begin de volgende vragen stellen:

- Sluit de informatie aan bij de doelstellingen van je onderzoek?
- Is de informatie waarschijnlijk nodig om een vraag te beantwoorden of conclusie te kunnen trekken?
- Is de informatie actueel (genoeg)?

- Eventueel: welk deel van de informatie is bruikbaar? Wil je bijvoorbeeld de conclusie van een rapport overnemen of alleen een deel van de bevindingen dat op die conclusie betrekking heeft?

Als op grond van deze vragen de informatie potentieel bruikbaar is in je onderzoek, ga dan verder met stap 2. Zo nee, gebruik de informatie dan niet.

Let op: leg in je onderzoeksdossier vast wat je conclusie is in deze stap en op grond van welke onderbouwing je daartoe bent gekomen. Als je de relevantie beperkt hebt getoetst, verantwoord dit dan in de nota van bevindingen.

Stap 2: Beoordeling kwaliteit: betrouwbaarheid en validiteit

De belangrijkste vraag die je jezelf bij deze stap moet stellen is of je zelf voldoende vertrouwen hebt dat de kwaliteit van de informatie toereikend is om te gebruiken en dat je dit vertrouwen ook overtuigend kunt onderbouwen en verantwoorden. Als je over het antwoord twijfelt of als het antwoord 'nee' is, dan is het advies: gebruik de informatie niet.

Bij de beoordeling van de kwaliteit van de informatie moet je je allereerst afvragen of de bron voldoende betrouwbaar is. Je kunt overwegen een uitgebreide beoordeling van de betrouwbaarheid en validiteit achterwege te laten wanneer je groot vertrouwen hebt in de bron. Bijvoorbeeld als de informatie afkomstig is uit:

- rapporten van zeer betrouwbare, onafhankelijke instituten, zoals CBS, RvS, WODC en planbureaus;
- artikelen in internationaal gereviewde tijdschriften;
- onafhankelijk onderzoek door een universiteit.

Dat betekent niet dat je al het onderzoek uit betrouwbaar geachte bron 'blind' kunt overnemen. Wees je er altijd van bewust dat het opnemen van derdeninformatie in een rapport zal worden opgevat als een keurmerk van de rekenkamer. Een lichte toets, bijvoorbeeld op plausibiliteit, is altijd aan te raden en kan soms aanleiding zijn om verder onderzoek te doen. Bovendien worden niet alle data van bijvoorbeeld het CBS op dezelfde manier verzameld en de betrouwbaarheid van deze gegevens kan dus ook verschillen. Ook planbureaus voeren studies uit met een variatie in diepgang. Informatie uit een *quick scan* heeft een andere mate van betrouwbaarheid dan informatie uit een jarenlang lopend onderzoek met een uitgewerkte en gecontroleerde methodiek. Universiteiten, planbureaus en andere onafhankelijke onderzoeksinstituten voeren niet alleen onafhankelijk onderzoek uit, maar ook onderzoek in opdracht. Wanneer het onderzoek in opdracht wordt uitgevoerd, is er goede reden om naar de opdrachtgever en diens belangen te kijken en op basis daarvan eventueel alsnog te besluiten de kwaliteit van de gegevens zelf te beoordelen door middel van de hiernavolgende stappen. Het is dus aan te raden niet alleen te kijken naar de betrouwbaarheid van het instituut als geheel maar ook naar

de achtergrond en reden voor het verzamelen en publiceren van de informatie die je zou willen gebruiken. Andere bronnen, denk bijvoorbeeld aan (websites van) uitvoeringsorganisaties van de overheid of van publieke en private organen, zijn niet per definitie onbetrouwbaar, maar als je gebruik wilt maken van die informatie, kun je niet zonder een kritische toets op betrouwbaarheid en validiteit. Dat geldt ook voor een belangrijke bron van informatie in onze onderzoeken, namelijk de *gecontroleerden*. De informatie uit bijvoorbeeld interviews en dossierstudie zal eveneens kritisch getoetst moeten worden!

Let op: leg in je onderzoeksdossier vast wat je conclusie is over de betrouwbaarheid van de bron en op grond van welke onderbouwing je daartoe bent gekomen. Als je de betrouwbaarheid van de bron beperkt hebt getoetst, verantwoord dit dan in de nota van bevindingen.

Betrouwbaarheid en validiteit zijn de belangrijkste kwaliteitseisen die je aan te gebruiken informatie kunt stellen. Voor een beoordeling van de betrouwbaarheid en validiteit van gegevens moet allereerst duidelijk zijn hoe cijfers, bevindingen en conclusies tot stand zijn gekomen. Je hebt kortom een goede beschrijving van het proces van dataverzameling en eventuele analyse nodig. Als deze ontbreekt kun je geen goed oordeel vellen over de betrouwbaarheid en de validiteit en moet je jezelf de vraag stellen of je deze informatie wel wilt gebruiken. Bij gebruik van open data is ook belangrijk of duidelijk is wat de oorspronkelijke bron van de data is.

De term *betrouwbaarheid* kan op twee manieren geïnterpreteerd worden. Ten eerste in de betekenis van nauwkeurigheid: de vraag of de meting nauwkeurig is toegepast en de gegevens niet te veel door toeval verstoord worden. De tweede betekenis gaat over de vraag in welke mate je erop kunt vertrouwen dat de gegeven informatie niet-vooringenomen en foutloos is. Beide interpretaties worden vaak door elkaar gebruikt. Eigenschappen die wijzen op een hoge mate van betrouwbaarheid in de zin van nauwkeurigheid zijn:

- een kleine spreiding in de meetgegevens;
- vergelijkbare uitkomsten bij gebruik van verschillende methoden;
- vergelijkbare uitkomsten bij meting op verschillende tijdstippen;
- vergelijkbare uitkomsten bij raadplegen van verschillende bronnen;
- vergelijkbare uitkomsten door verschillende onderzoekers;
- een bevredigende beschrijving van de manier waarop toevalsfactoren zijn uitgesloten;
- bij statistische analyse: of deze goed is uitgevoerd en de resultaten juist worden geïnterpreteerd. Is er sprake van een acceptabele betrouwbaarheids-/onzekerheidsinterval?

De laatste drie *bullets* kunnen ook een indicatie geven voor het meer subjectieve waardeoordeel over de betrouwbaarheid. Daarnaast kun je nog een aantal andere indicaties gebruiken zoals:

- Gegevens afkomstig van een onafhankelijke bron zonder duidelijk belang bij de uitkomst (universiteit, planbureau, als geen sprake is van onderzoek in opdracht) zijn vaak betrouwbaarder dan gegevens verzameld in opdracht.
- De mogelijke belangen van de opdrachtgever bij onderzoek in opdracht en de mogelijkheid van kleuring van de informatie.
- Informatie die geaccepteerd wordt door partijen met duidelijk tegengestelde belangen is waarschijnlijk betrouwbaarder dan informatie waar dat niet het geval is (anders gezegd: beleidsinformatie verkregen van Shell over maatschappelijk verantwoord ondernemen kun je natuurlijk toetsen aan informatie van de werkgroep Duurzaam ondernemen bij VNO/NCW, maar sterker is een toets aan informatie van Greenpeace of Milieudefensie).
- Wanneer je slechts informatie uit één bron gebruikt die ook nog een bepaald belang vertegenwoordigt, is toetsing extra belangrijk. Een simpele eerste toets is om via Google te zoeken of er ook andere informatie of andere meningen te vinden zijn.
- Ziet de informatie er slordig uit, zitten er fouten of onlogische waarden in? Anderzijds kan informatie er ook te netjes uitzien. Tabellen zonder *missing values* kunnen soms ook verdacht zijn.
- Informatie die is verkregen door rechtstreekse meting of analyse is betrouwbaarder dan indirect verkregen informatie.
- De betrouwbaarheid van gegevens kan soms indicatief worden afgeleid uit het systeem van kwaliteitscontrole.
- Soms kun je leunen op de zwaarte van het juridisch formele proces waarmee de gegevens tot stand zijn gekomen. Informatie in een wet is waarschijnlijk betrouwbaarder dan in een beschikking of een circulaire. Informatie in bijlagen bij formele verantwoordingsstukken of in Kamerstukken zouden betrouwbaarder moeten zijn dan memo's van beleidsdirecties.
- Wanneer je niet zeker bent van de betrouwbaarheid van je gegevens kan je ook navragen bij deskundigen en stakeholders of zij deze informatie betrouwbaar vinden. Wees je daarbij uiteraard wel bewust van de belangen van de stakeholder.
- Bij geografische data: is helder welke grenzen zijn gebruikt bij het afbakenen van geografische eenheden? Hoe zijn de data verzameld, met welk doel? Zijn de data juist, volledig en consistent?

Validiteit richt zich op de vraag of de gegevens inhoudelijk en adequaat gemeten zijn.

Voor het beoordelen van validiteit kun je de volgende vragen stellen:

- Zijn begrippen zodanig geoperationaliseerd in meetbare variabelen dat je meet wat je meten wilt (begripsvaliditeit)?
- Worden de variabelen op een goede manier en met de juiste methoden gemeten (inhoudsvaliditeit)?
- In welke mate kunnen causale conclusies worden getrokken uit de metingen (interne validiteit)? Bij interne validiteit gaat het over de vraag of gevonden effecten een causaal gevolg zijn van de beleidsingreep.
- In hoeverre zijn de conclusies geldig buiten de actueel onderzochte eenheden (externe validiteit, generaliseerbaarheid)?

Hierna valt dan de beslissing wel of niet gebruiken, of wellicht gebruiken onder voorwaarden. Deze beslissing hangt, naast de kwaliteit van de informatie zelf, ook af van het doel waarvoor je deze informatie wilt gebruiken en van het type informatie (bijvoorbeeld meetgegevens, verslagen of conclusies uit een onderzoek).

Bedenk allereerst of je de informatie wilt gebruiken ter onderbouwing van conclusies of als illustratie. Als je de informatie wilt gebruiken ter onderbouwing van je conclusies, zal je hogere eisen stellen aan de validiteit, betrouwbaarheid en relevantie dan als je een deel van de informatie als voorbeeld wilt gebruiken.

Bijvoorbeeld om de mening van een stakeholder te illustreren.

Ook het type informatie kan het oordeel over wel of niet gebruiken beïnvloeden.

Meetgegevens en monitorinformatie kun je op een andere manier gebruiken en deze zal je daarom waarschijnlijk ook anders beoordelen dan bijvoorbeeld een onderzoeksrapport dat een uitspraak doet over de effectiviteit of doelmatigheid van beleid. Wanneer de conclusie niet onderbouwd wordt door de achterliggende bevindingen, maar deze bevindingen op zichzelf wel valide en betrouwbaar zijn kan dat een reden zijn om alleen dat deel van een rapport te gebruiken.

Let op: leg in je onderzoekdossier vast wat je conclusie is in deze stap en op grond van welke onderbouwing je daartoe bent gekomen. Als je de betrouwbaarheid en validiteit beperkt hebt getoetst, verantwoord dit dan in de nota van bevindingen. Wees je er ook van bewust dat de mate van toetsing bepalend kan zijn voor de robuustheid van de conclusies!

Voorbeeld: In het onderzoek [Autobelastingen als beleidsinstrument](#) is gebruik gemaakt van gegevens over de typekeuring van auto's die is opgesteld door de RDW. Het team heeft die informatie niet op kwaliteit getoetst en rapporteert dat ook transparant in de methodologische bijlage. Hoe zou je deze informatie eventueel kunnen toetsen? Een paar opties:

- De typekeuring overdoen. Maar dat is geen realistische optie. Als we dat voor alle informatie die we benutten zouden moeten doen, zouden we niet meer aan ander onderzoek toekomen.

- Nagaan of de RDW een systeem waarmee de medewerkers zelf de juistheid en volledigheid van de informatie in het systeem kunnen bepalen. En nagaan of daarover rapportages zijn en of de uitkomsten daarvan aangeven dat die informatie juist en volledig is. Dit lijkt een uitvoerbare stap, althans voor databestanden. Tegelijkertijd: als je dat voor alle gebruikte databestanden doet, vergt dit wel extra tijd, werk en doorlooptijd.
- Een lichte en minder tijdrovende variant van de vorige optie zou kunnen zijn om aan de betrokken organisatie te vragen of en hoe ze de betrouwbaarheid en validiteit gewaarborgd hebben en – mits dat antwoord overtuigend en vertrouwenwekkend klinkt – hierop te steunen.
- Professional judgement: afgaan op je professionaliteit, kennis van het veld en je wetenschappelijke achtergrond. Dit moet je dan wel aannemelijk maken in je onderzoeksverantwoording.
- Afgaan op de plausibiliteit van de uitkomsten van de data. Dit kun je doen door goed naar de data te kijken en je gezond verstand te gebruiken, er logische checks op uit te voeren en aan experts te vragen of zij deze datasets bruikbaar vinden. Dat klinkt mooi, het zal tot op zekere hoogte ook wel mogelijk zijn, maar alle fouten zul je er niet mee achterhalen. Dat hoeft op zichzelf geen probleem te zijn. In het onderzoek Autobelastingen leiden alle berekeningen tot punt-schattingen: getallen op de euro en ton CO₂ nauwkeurig. We hebben als Algemene Rekenkamer geen idee of die getallen kloppen, maar voor de algemene conclusie maakt het niet uit.
- Afgaan op de authenticiteit van de bron. De RDW is een betrouwbare organisatie (dat vermoeden we, we hebben geen bewijs). Je zou daartegenin kunnen brengen, dat er zoveel problemen bij de RDW spelen dat dit ook wel eens negatieve gevolgen voor de kwaliteit van hun data zou kunnen hebben.

Zo moet je in elk concreet onderzoek afwegen wat zinvol en haalbaar is. Maar, zoals gezegd, elke keuze heeft gevolgen voor de robuustheid van conclusies.

8.6.3 Gebruik van derdeninformatie

In 08.06.02 hebben we ons ervan vergewist dat de derdeninformatie die we willen gebruiken (potentieel) relevant en kwalitatief voldoende is. Nu moeten we deze informatie uiteraard ook nog zelf goed gebruiken (*noblesse oblige*).

We kunnen globaal de volgende criteria onderscheiden waaraan ons eigen gebruik zou moeten voldoen:

- Relevant
- Evenwichtig
- Transparant

Als je besloten hebt de informatie van derden te gebruiken, moet je nagaan of de informatie waarvan je het idee had dat deze potentieel relevant was voor je onderzoek (zie stap 1), dat ook echt is voor het doel waarvoor je deze wilt gebruiken in je rapport (of ander product). Dit kan bijvoorbeeld zijn: het onderbouwen van een conclusie, beschrijven van een situatie of geven van een illustratie. Je kijkt nu specifieker dan in stap 1, toen je waarschijnlijk nog niet duidelijk voor ogen had welke boodschap of conclusies je wilde overbrengen. Je kunt jezelf de volgende vragen stellen:

- Heeft de gebruikte informatie een logische relatie met de uitspraak of eventuele conclusie die erop wordt gebaseerd?
- Volledigheid: is alle informatie die, gelet op het doel van het gebruik, beschikbaar zou moeten zijn ook beschikbaar? Is het bewijs overtuigend?

Evenwichtigheid richt zich op de vraag of de informatie eerlijk, objectief en gebalanceerd wordt gebruikt. Heb je alle beschikbare relevante informatie gewogen en deze naar verhouding van het gewicht van de informatie gebruikt? Wanneer je bijvoorbeeld slechts een deel van de conclusies uit een onderzoek betreft in je onderzoek, wees je er dan van bewust waarom je de andere conclusies niet meeneemt. Is dat omdat ze niet relevant zijn voor je eigen conclusies, of omdat ze niet prettig aansluiten bij je eigen redeneerlijn? Wanneer je voor de onderbouwing van een conclusie slechts beschikt over informatie van één belanghebbende partij, moet je je afvragen of je de conclusie kunt trekken zonder deze informatie te toetsen bij andere stakeholders. Zie ook de tips bij betrouwbaarheid van informatie.

Transparantie geeft inzicht in de kwaliteit van de gebruikte gegevens en de herkomst ervan. Geef ook helder aan hoe de derdeninformatie is gebruikt en welke beperkingen er eventueel aan kleven. Soms zijn conclusies in een extern rapport te positief opgeschreven in relatie tot de bevindingen van het onderzoek, maar zijn de bevindingen op zichzelf betrouwbaar en valide. Als je dit inzichtelijk maakt, kun je de relevante, betrouwbare en valide delen gebruiken bij je eigen onderzoek. Een speciaal punt van aandacht hierbij zijn toekomstscenario's en modelberekeningen. Deze zijn vrijwel altijd gebaseerd op een aantal aannames en vereenvoudigingen van de werkelijkheid. Van belang is dat die aannames expliciet gemaakt worden. Wanneer relevante aspecten of factoren buiten beschouwing worden gelaten, moet dit expliciet zijn vermeld en gemotiveerd.

8.7 Normen voor de beoordeling van de overheidsrol in beleidsnetwerken

We spreken van een beleidsnetwerk wanneer een minister zijn of haar beleid op een bepaald terrein wil uitwerken samen met andere overheden, ondernemingen, maatschappelijke organisaties en/of groepen burgers. Dat 'samen uitwerken' heeft veelal niet de vorm van een bescheiden consultatieronde, maar van langdurige samenwerkingsverbanden die zowel de voorbereiding als de uitvoering van beleid kunnen omvatten – vandaar de term 'beleidsnetwerk'.

Norm: de noodzaak tot netwerksturing moet onderbouwd zijn

Sturing door het vormen en ontwikkelen van een netwerk is niet per se 'beter' dan hiërarchische sturing. Een eerste norm die men aan overheidshandelen kan stellen is dat goed onderbouwd moet zijn waarom ze in een bepaald geval kiest voor netwerksturing, of waarom dit de enige mogelijkheid is. Dit kan het geval zijn wanneer grote onzekerheid heerst over de definitie, de oorzaken en de oplossingen van het probleem waarvoor beleid wordt gemaakt. Eenzijdig bepaalde maatregelen opleggen aan betrokken actoren heeft in zulke situaties geen zin; dit zal door een gebrek aan draagvlak niet werken. Behalve onzekerheid kunnen er nog andere redenen zijn waarom de overheid soms wel moet samenwerken met andere actoren:

- anderen kunnen verhinderen dat op het desbetreffende gebied beleid van de grond komt (ze hebben 'hidermacht');
- anderen hebben specifieke kennis die onontbeerlijk is;
- anderen hebben geld, de overheid zelf niet of niet genoeg.

Norm: doelen moeten tijdig en SMART-C worden geformuleerd

In netwerkachtige situaties worden doelen veelal niet vooraf vastgelegd; ze komen in interactie met derden tot stand. Bovendien kunnen nieuwe inzichten en ervaringen leiden tot bijstellingen. Tegen die achtergrond kunnen de volgende normen worden gehanteerd:

- In gevallen waar sprake is van netwerksturing hoeven de doelstellingen niet in een vroeg stadium nauwkeurig te worden vastgesteld. Wel mag een fasering van het proces worden verwacht.
- Deze fasering bevat een moment waarop de doelstellingen wel (evalueerbaar, SMART-C) geformuleerd moeten zijn. Dit moment ligt voor de start van de uitvoering van het project, dat wil zeggen voordat, afgezien van proces- en apparaatskosten, daadwerkelijk uitgaven worden gedaan.

- Beleidsdoelen kunnen om verschillende redenen in de loop van het proces wijzigen. Wanneer dit gebeurt dient dit goed te worden onderbouwd en verantwoord. De gewijzigde doelen dienen opnieuw aan de normen van evalueerbaarheid te voldoen.

Norm: voor (tussen)producten moeten evaluatiecriteria worden geformuleerd

Voor netwerksturing kunnen verschillende uitkomsten worden onderscheiden. In de eerste plaats kan een overeenkomst (bijvoorbeeld een bestuursovereenkomst) als uitkomst (= prestatie) worden gezien. In vervolg hierop kan ook de uitvoering van een in een overeenkomst afgesproken beleidsprogramma als uitkomst worden gezien. Deze verschillende producten kunnen afzonderlijk worden beoordeeld aan de hand van de volgende vragen:

In welke mate is er een deugdelijke relatie tussen de voorgestelde of ingezette instrumenten en het te bereiken doel?

- In welke mate wordt de uitkomst door relevante deelnemende actoren geaccepteerd?
- Tegen welke kosten is het resultaat tot stand gekomen?
- Binnen welke tijd is het resultaat tot stand gekomen?

Aan het eind van de uitvoering van het beleidsprogramma geldt vanzelfsprekend als extra evaluatiecriterium de vraag of met de geleverde prestaties de beoogde effecten zijn behaald (doeltreffendheid) en in welke verhouding dit staat tot de gemaakte kosten (doelmatigheid).

Norm: de overheid moet haar rol in beleidsnetwerken doeltreffend vervullen

De overheid kan in een beleidsnetwerk verschillende rollen vervullen. Zij kan een actor zijn te midden van andere deelnemers aan het netwerk en 'het spel moeten meespelen'. Doorgaans zal ze echter als regisseur optreden (dan wel een procesmanager aanstellen).

De overheid als medespeler in een beleidsnetwerk

Om de rol van medespeler goed te vervullen, moet de overheid:

- belangen, macht en mogelijkheden van de diverse actoren goed in het oog hebben;
- zorgen dat ze 'wisselgeld' achter de hand heeft en dit ook pas 'uitgeven' als het echt nodig is;
- een betrouwbare en prudente partner zijn om het gezag als overheid niet op het spel te zetten en om het vertrouwen van de andere spelers niet te verliezen.

De overheid als regisseur in een beleidsnetwerk

Om de rol van regisseur goed te vervullen, moet de overheid, respectievelijk de procesmanager:

- een stakeholderanalyse uitvoeren, dat wil zeggen nagaan wie om welke reden bij het netwerk betrokken zou moeten worden;
- zorgen voor een duidelijk 'binding', een afspraak over wat men wil bereiken en de wijze waarop men wil samenwerken;
- zorgen voor helderheid over het mandaat van de deelnemende partijen, en over de omvang van zowel de proces- als de programma-uitgaven;
- zorgen dat de randvoorwaarden voor samenwerking expliciet vastliggen, bijvoorbeeld helder maken welke uitgangspunten (van beleid) niet ter discussie staan;
- zorgen dat de procesafspraken zijn vastgelegd, onder andere: wie verantwoordelijk is voor het management, hoe de openheid wordt gegarandeerd;
- helderheid verschaffen over de vraag welke inhoudelijke kennis wordt geaccepteerd door alle betrokken partijen, en welke feiten worden betwist. Dat wil zeggen dat de overheid een inspanningsverplichting om de gemeenschappelijk geaccepteerde kennis te vergroten;
- helder communiceren tijdens het gehele proces, onder andere door tussendoelen en wijzigingen van doelstellingen altijd expliciet te maken en te zorgen dat deze aan alle actoren duidelijk zijn.

Norm: de overheid moet verantwoording afleggen over haar rol in beleidsnetwerken

Wanneer de overheid netwerksturing toepast dient zij een aantal regels in acht te nemen met het oog op de democratische verantwoording die zij moet afleggen.

- Hoewel in een beleidsnetwerk het proces van besluitvorming en uitvoering niet lineair en rationeel verloopt, dient de overheid wel op gezette tijden verantwoording af te leggen over de wijze waarop zij het belang van de overheid in het netwerk heeft behartigd. De verantwoordelijke minister dient met de Tweede Kamer af te spreken waarover en op welk moment de Tweede Kamer wordt geïnformeerd.
- Als de overheid de regie heeft gevoerd of een procesmanager heeft aangesteld, is zij (doorgaans één minister) direct verantwoordelijk voor de kwaliteit van de regie. Daarover moet dus in elk geval verantwoording worden afgelegd aan de Tweede Kamer.

8.8 Normen voor toezicht op naleving en handhaving van EU wet- en regelgeving

De normen waaraan de rekenkamer het nalevingstoezicht en de handhaving van EU-wet- en -regelgeving toetst, vinden hun grondslag in de EU-verdragen en de daarop gebaseerde secundaire regelgeving (zoals EU-richtlijnen en -verordeningen), en de jurisprudentie van het Europese Hof van Justitie. Hieraan zijn alle EU-lidstaten gebonden. Binnen de EU ligt de verantwoordelijkheid voor het toezicht op de naleving van EU-wet- en -regelgeving primair bij de lidstaten. Hoe het toezicht en de handhaving worden vormgegeven, wordt echter in belangrijke mate op EU-niveau bepaald.

De volgende normen zijn relevant bij het onderzoek naar toezicht op naleving en handhaving van EU-wet- en -regelgeving.

1. *De lidstaat draagt zorg dat de Nederlandse rechtsorde verenigbaar is met de regels en beginselen van het EU-recht.* In de Europese rechtsorde worden de gedragingen van burgers, instellingen en lidstaten getoetst op hun verenigbaarheid met de regels en beginselen van het communautair recht. De EU-wetgever is voor de omzetting, uitvoering, toezicht en handhaving van de meeste EU-wet- en -regelgeving afhankelijk van de nationale instanties. De effectiviteit van de EU-wet- en -regelgeving hangt zodoende in grote mate af van de werking van de nationale rechtsstelsels.
2. *Ministers moeten al het mogelijke doen om doeltreffende uitvoering te verzekeren van het EU-beleid en moeten zorg dragen voor juiste besteding en goed beheer van de gelden voor EU-programma's die Nederland gezamenlijk met de Europese Commissie uitvoert.* Veel EU-programma's worden uitgevoerd door decentrale bestuursorganen waarover de minister geen directe zeggenschap heeft, terwijl de lidstaat als geheel verantwoordelijk is, als verplichtingen niet worden nagekomen. Daarom is een specifiek toezichtstelsel in het leven geroepen, waardoor de minister aanwijzingen kan geven aan het bestuursorgaan over de rechtmatige en de doelmatige aanwending van de EU-subsidies. De minister kan bedragen verhalen op het bestuursorgaan als door diens handelen of nalaten de lidstaat door het Europese Hof van Justitie verplicht wordt tot het betalen van een forfaitaire som of dwangsom.
3. *Bij nieuwe Europese richtlijnen en verordeningen moet de verantwoordelijke minister nationale instanties aanwijzen die met de uitvoering en handhaving worden belast. Daarnaast moeten ministers waarborgen dat de handhavende activiteiten van deze instanties effectief en proportioneel zijn, niet in strijd zijn met de voorwaarden van het nationale recht en een afschrikwekkend effect hebben.* De

EU heeft bepaald dat lidstaten overtredingen van EU-wet- en -regelgeving niet mogen gedogen. Tegen geconstateerde overtredingen moet dus altijd sanctionerend (bestuursrechtelijk of strafrechtelijk) worden opgetreden.

4. *Ministers moeten zorg dragen voor een tijdige, juiste en volledige omzetting van EU-richtlijnen in nationale wet- en regelgeving, en voor directe uitvoering en handhaving van EU-verordeningen.* Een lidstaat die de voorgeschreven termijn voor de omzetting van richtlijnen overschrijdt, kan zich niet beroepen op zijn interne rechtsorde. Voor verordeningen geldt dat zij direct rechten en plichten met zich meebrengen voor burgers, bedrijven, lidstaten of Europese instellingen. De EU schrijft voor dat lidstaten zodanige maatregelen treffen dat verordeningen direct in de nationale rechtsorde kunnen worden uitgevoerd en gehandhaafd; lidstaten dienen daar ook op te anticiperen.
5. *Nederlandse toezichthouders dienen waar relevant met andere toezichthouders samen te werken.* Europees en nationaal toezicht zijn nauw met elkaar verweven. De Europese Commissie en nationale toezichthouders moeten daarom zo veel mogelijk samenwerken om consistente toepassing van het EU-recht te bevorderen.
6. *De lidstaat dient medewerking te verlenen aan toezichtactiviteiten van de Europese Commissie.* De Europese Commissie ziet erop toe dat lidstaten het EU-recht toepassen en de handhavingsplicht vervullen. Het toezicht van de Commissie kent twee hoofdvormen: bestuurlijk toezicht (controle op hoe de EU-instellingen en de lidstaten – en de daarbinnen werkzame overheidsinstanties – het EU-recht toepassen en handhaven) en nalevingstoezicht (zelfstandige, operationele en rechtstreekse controle op de naleving van het EU-recht in de lidstaten).

Bijlage: uitwerking beoordelen van beleidsdoelen

Beleidsdoelen moeten SMART-C geformuleerd zijn: specifiek, meetbaar, afgestemd, realistisch, tijdgebonden en consistent.

Specifiek: Doelstellingen moeten concreet worden geformuleerd: welke prestaties en effecten wil men realiseren en welke doelgroepen (zo die er zijn) wil men bereiken?

Toelichting

Doelen moeten bij voorkeur geformuleerd worden in termen van *prestaties* en te bereiken maatschappelijke *effecten*.

Een prestatiedoelstelling heeft betrekking op een beleidsprestatie die door de gecontroleerde kan worden gerealiseerd, zonder dat deze daarbij afhankelijk is van anderen of van maatschappelijke ontwikkelingen die buiten de invloed van het beleid liggen. Bij een doelstelling gericht op maatschappelijk effect ligt dat anders. Er zijn naast het beleid meestal nog meer factoren die van invloed zijn op het bereiken van de desbetreffende doelstelling. Wanneer het niet mogelijk is een doelstelling te formuleren die gericht is op een prestatie of een maatschappelijk effect, kan worden teruggevallen op een procesdoelstelling. Denk bijvoorbeeld aan productkwaliteit die vaak moeilijk meetbaar te maken is. Het doel kan dan zijn: de voortgang van maatregelen ter verbetering van de kwaliteit (proces).

Als het beleid niet op de hele bevolking is gericht, moet de doelgroep van het beleid zijn omschreven en zo duidelijk mogelijk zijn afgebakend. Ook moet de keuze van de doelgroep zijn gemotiveerd.

Meetbaar: Uiteindelijk moet vast te stellen zijn of de doelstellingen zijn gehaald. Hiervoor kan een nulmeting nodig zijn. Eenduidige definiëring van begrippen is vereist.

Toelichting

Gemeten moet kunnen worden, aan de hand van concrete indicatoren, of beleidsdoelen al dan niet bereikt zijn. Beleidsdoelen moeten toetsbaar zijn geformuleerd. De beoogde situatie bij doelrealisatie moet eenduidig zijn omschreven, (definities van) begrippen moeten slechts voor een uitleg vatbaar zijn en meetbaar zijn gedefinieerd, zodat duidelijk is waarop de minister kan worden aangesproken.

Bij doelen ten opzichte van een referentiepunt moet de situatie van het referentiepunt bekend zijn. Een nulmeting is daarvoor niet altijd noodzakelijk. Voorbeeld: de doelstelling "Het aantal geluidgehinderden mag niet groter zijn dan in 2014", vereist

een nulmeting van de situatie in 2014. De doelstelling “Het aantal geluidgehinderden mag in 2020 niet groter zijn dan 12.000” niet.

Afgestemd: Doelstellingen moeten expliciet zijn gemaakt in Tweede Kamerstukken of beleidsdocumenten.

Toelichting

Beleidsdoelen moeten zijn afgestemd met de Tweede Kamer en met de uitvoerders van het beleid. Ook tussentijdse bijstellingen of wijzigingen van de doelstelling moeten met hen worden afgestemd.

Voor organisaties die publieke taken uitvoeren is deze norm eveneens van toepassing, maar spelen ook andere actoren een rol. Bij deze organisaties is maatschappelijk draagvlak voor het beleid belangrijk. Draagvlak krijgt een organisatie door goed te luisteren naar de problemen en behoeften van burgers en door de dialoog met de maatschappelijke omgeving aan te gaan. Deze organisaties formuleren op basis daarvan – vanuit de eigen verantwoordelijkheid – beleidsdoelen, en zijn transparant over de afwegingen die zij daarbij maken.

Let op: in het IAK staat de A voor Acceptabel (doelen moeten zodanig worden geformuleerd dat ze aanvaard, begrepen en eenduidig uitgelegd zullen worden binnen de doelgroep die ze moeten gaan uitvoeren). Het begrip Afgestemd is breder en omvat bijvoorbeeld ook de communicatie met de Tweede Kamer..

Realistisch: Doelstellingen moeten zodanig zijn gekozen dat ze met het beschikbare budget bereikt kunnen worden, rekening houdend met plausibele scenario's voor de invloed van externe factoren en beleidsmaatregelen.

Toelichting

Beleidsdoelen moeten zodanig zijn gekozen dat ze realiseerbaar zijn onder alle plausibele omstandigheden (exogene variabelen en beleidsmaatregelen).

Het realiteitsgehalte van beleidsdoelen valt uiteen in twee elementen:

1. De ambities moeten op een haalbaar (maar voldoende uitdagend) niveau zijn geformuleerd.
2. Er moet bij de doelformulering rekening zijn gehouden met eventuele beperkte mogelijkheden om de doelen te behalen (exogene factoren en gedeelde beleids- of resultaatverantwoordelijkheden).

Tijdgebonden: Er moet een datum zijn vastgelegd waarop de einddoelstelling moet zijn bereikt. Bij langlopend beleid kunnen ook tussendoelstellingen van belang zijn om tussentijds vast te stellen of men op het juiste spoor zit. In dat geval moet ook

een tijdpad met gespecificeerde mijlpalen zijn aangegeven (bijvoorbeeld tussendoelen en bijbehorende data, of het tempo waarin het einddoel moet zijn gerealiseerd).

Consistent: De doelstellingen moeten onderling consistent zijn (het ene doel moet het andere niet uitsluiten) en consistent met de basisgegevens (als de basisgegevens veranderen, moeten de doelen zo nodig ook worden aangepast).

Toelichting

1. Consistentie tussen doelen:

- Beleidsdoelen moeten met elkaar te verenigen zijn.
- (Tussen)doelen voor onderdelen van het beleid moeten aansluiten bij de algemene doelen voor het gehele beleidsterrein (hiërarchische ordening).
- Bij decentraal uitgevoerd beleid moeten de doelen van de rijksoverheid en die van de uitvoerende instanties (bijvoorbeeld gemeenten, uitvoerende instellingen) op elkaar blijven aansluiten.
- De formulering van de doelstelling moet in de tijd consistent zijn.

Vaak zijn aan het beleid verscheidene doelen verbonden. Doelen kunnen geordend worden in een 'doelenboom': in een hiërarchie (doelen per maatregel, doelen voor onderdelen van het beleid en algemene doelen voor het gehele beleidsterrein) of in de tijd (tussendoelen, einddoel). Van belang is dat deze doelen consistent met elkaar zijn. Het bereiken van het ene doel moet niet op gespannen voet staan met of belemmerend zijn voor het bereiken van het andere doel.

2. Consistentie tussen doelen en basisgegevens:

Doelstellingen moeten aansluiten op de basisgegevens waarop ze gebaseerd zijn. Als deze gegevens veranderen, moeten de doelstellingen worden aangepast.

Voorbeeld: de doelstelling om de hoeveelheid afval tussen 2015 en 2025 maximaal te laten groeien van 40 naar 50 miljoen ton is gebaseerd op de basisinformatie dat de hoeveelheid afval in 2015 40 miljoen ton bedroeg. Als later blijkt dat er in 2015 geen 40 maar 50 miljoen ton afval was, is de doelstelling niet meer consistent met de informatie waarop deze gebaseerd is.

9. Methoden en technieken van dataverzameling en -analyse

9.1 Inleiding

In de hoofdstukken over beschrijvende, toetsende en verklarende onderzoeksvragen is kort ingegaan op mogelijke bruikbare methoden en technieken bij de dataverzameling en de analyse van data. In dit hoofdstuk gaan we dieper in op de verschillende methoden en technieken die je kunt gebruiken in doelmatigheids- en doeltreffendheidsonderzoek. We staan hierbij apart stil bij dataverzameling (paragraaf 9.2) en data-analyse (paragraaf 9.3).

9.2 Methoden en technieken van dataverzameling

Data (in het kader van rekenkameronderzoek ook wel bewijsmateriaal of auditbewijs genoemd) voor het beantwoorden van onderzoeksvragen worden in 3 categorieën verdeeld:

- bewijs uit interviews (9.2.1);
- bewijs uit documenten (9.2.2);
- fysiek bewijs (9.2.3).

Je kunt veel verschillende soorten (kwantitatieve of kwalitatieve) informatie verzamelen. Het belangrijkste is dat de data iets zeggen over de eenheden in je onderzoek en dat ze nodig zijn om de onderzoeksvraag te beantwoorden. Als je de effecten van beleid wilt vaststellen (bijvoorbeeld: in welke mate is het beleid doeltreffend geweest?), heb je hiervoor vaak kwantitatieve gegevens nodig. Deze gegevens moeten aan bepaalde eisen voldoen. In je onderzoeksdesign moet je hiermee goed rekening houden.

Vaak kun je met kwalitatieve gegevens de 'waaromvraag' beantwoorden.

Bijvoorbeeld: waarom werkt het beleid niet zoals beoogd? Met kwalitatieve informatie kun je bijvoorbeeld motieven, achterliggende oorzaken en redeneerlijnen (beleidstheorie) onderzoeken. In veel gevallen is het zinvol om te werken met een combinatie van kwantitatieve en kwalitatieve gegevens.

9.2.1 Bewijs uit interviews

Bij bewijs uit interviews maken we onderscheid tussen individuele interviews, groepsinterviews en enquêtes (of vragenlijsten). Vooral bij die laatste vorm van dataverzameling is steekproefselectie van groot belang. Deze onderwerpen komen hieronder aan bod.

9.2.1.1 Individueel interview

Met een individueel interview kan de mening en kennis van één persoon uitgediept worden. Diepte-interviews worden in onderzoek vaak gebruikt om bijvoorbeeld meer kennis op te doen over een bepaald onderwerp of informatie boven tafel te krijgen. Een interview kent in het onderzoek drie fasen: (1) de voorbereiding, (2) het afnemen van het interview en (3) het verwerken van de resultaten.

1. Voorbereiden interview

Voordat het interview plaatsvindt, is het nuttig om na te denken over de verschillende manieren waarop een interview afgenomen kan worden. Veel interviews vinden plaats aan de hand van een topiclijst. Deze lijst kan bijvoorbeeld op basis van de onderzoeksvragen van het project worden samengesteld. Op deze lijst staan de onderwerpen die de interviewer graag wil bespreken met zijn respondent. Het gebruik van een topiclijst geeft richting aan het interview, zorgt ervoor dat je geen onderwerpen vergeet die belangrijk zijn voor het onderzoek, maar geeft ook ruimte om op het ene onderwerp dieper in te gaan dan op het andere. Meer gestructureerd is een lijst met concrete vragen die de interviewer wil stellen. Het interview kent dan minder spontaniteit maar het zorgt er wel voor dat alle vragen behandeld worden in het interview. Dit kan nodig zijn wanneer je een zekere mate van standaardisatie wilt, bijvoorbeeld als je meer interviews afneemt en deze ook statistisch of anderszins systematisch wilt vergelijken.

Er zijn nog meer technieken voor het afnemen van interviews. Zo kunnen er stellingen aan de respondenten worden voorgelegd, waarop zij kunnen reageren. Dit lokt vaak een andere reactie uit dan wanneer er een open vraag over hetzelfde onderwerp wordt gesteld. Ook kan er gebruikgemaakt worden van de kritische-incidentenmethode (KIM). Hierbij wordt de respondent gevraagd een bepaalde gebeurtenis, incident of ervaring uit zijn/haar werkpraktijk op te halen en te vertellen.

Aan de hand van die gebeurtenis vraagt de interviewer dan verder door. Dit is een handige methode als het doel is om bepaalde processen te analyseren. Dit kan effectiever zijn dan het stellen van directe vragen.

In de context van een (open) interview kan een deel ook meer gestructureerd uitgevraagd worden in de vorm van een vignet, een korte beschrijving van een bepaalde situatie of casus. Daarbij wordt de respondent gevraagd aan te geven hoe hij/zij in een bepaald scenario zou oordelen en handelen. Zo zou je 1 of 2 vignetten kunnen voorleggen aan respondenten om te bekijken hoe verschillende mensen dezelfde situatie beoordelen. Denk bijvoorbeeld aan een casus met een potentiële milieuovertreding, waarbij aan verschillende toezichthouders wordt gevraagd wat naar hun mening de meest passende aanpak is in die concrete casus. Dit is een voorbeeld van een vereenvoudigde versie van de vignettenmethode.

2. Afnemen interview

Ook voor het afnemen van het interviews zijn bepaalde technieken, met name gespreksvaardigheden, van belang. Belangrijk is dat je duidelijk maakt wat het doel is van het interview en in welke context dit plaatsvindt. Benoem de verschillende rollen, maar benadruk ook het gemeenschappelijk belang (verbeteren van het functioneren). Let erop dat je vragen ook echt worden beantwoord. Stel zo nodig aanvullende vragen of vraag hetzelfde in andere woorden. Wees je ook bewust van mechanismen die jouw oordeelsvorming beïnvloeden. Als je interviews gebruikt om de bruikbaarheid van informatie te bepalen, wat bijvoorbeeld bij ministeries vaak zal voorkomen, dan is het mogelijk dat je gesprekspartner je beïnvloedt met zijn of haar charmes, autoriteit, lichaamstaal (zuchten) en houding. Stel, iemand maakt duidelijk dat hij weinig tijd heeft en raffelt het interview af. Kom je dan terug of neem je genoegen met de summiere informatie, zodat je deze zielige, want extreem drukke persoon niet nog een keer lastig moet vallen? Wat ook kan meespelen is de relatie met de persoon, die je om wat voor reden dan ook goed wilt houden. Alle genoemde aspecten kunnen maken dat je genoegen neemt met onvolledige antwoorden, niet doorvraagt en geen nieuwe vragen stelt. Dat kan tot gevolg hebben dat je onvoldoende basis hebt om op de informatie te kunnen steunen.

3. Verwerken resultaten interview

Er zijn verschillende mogelijkheden om de resultaten van interviews op een kwalitatieve manier te verwerken. Je kunt quotes gebruiken om bepaalde bevindingen te illustreren. Maar ook samenvattingen van verschillende interviews gebruiken om een beeld te geven van de verschillen of overeenkomsten van antwoorden die deelnemers op een bepaalde vraag gaven: zien alle geïnterviewden bijvoorbeeld hetzelfde probleem of komen ze met verschillende inzichten? Door

middel van tekstanalyse kun je deze verwerking ook systematisch aanpakken. Dit kan bijvoorbeeld door met tellingen te laten zien hoe vaak een bepaald onderwerp wordt genoemd. Daarvoor is het nodig om de interviews op te nemen en letterlijk uit te schrijven (transcriberen). Vervolgens kunnen tekstpassages aan de hand van een codeerschema ondergebracht worden in categorieën (coderen). Dan kan geanalyseerd worden hoe vaak een bepaalde code voorkomt. Een handige tool hiervoor is Atlas Ti. Wanneer gesloten vragen zijn gesteld in een gestructureerd interview kunnen de resultaten ook kwantitatief worden verwerkt. .

9.2.1.2 Groepsinterview

Een groepsinterview of focusgroep kan uitgebreide en diepgaande informatie opleveren over een bepaald onderwerp door de interactie tussen een aantal deelnemers met kennis over het onderwerp.

Focusgroepen zijn bijeenkomsten voor ongeveer 5 tot 20 mensen, bijvoorbeeld vertegenwoordigers van verschillende veldorganisaties waar het onderzoek zich op richt. In de bijeenkomst wordt een specifiek onderwerp behandeld, waarover door de deelnemers gediscussieerd wordt. Deze discussie wordt geobserveerd door de organisator en vastgelegd. Een focusgroep is daarmee een kruising tussen een discussie en een interview. Het doel van de focusgroep is om meningen en ideeën van verschillende mensen over een bepaald onderwerp te vergaren en daar dieper op in te kunnen gaan. Door het observeren van de discussie tussen verschillende belanghebbenden hoor je wellicht andere dingen dan in een een-op-een interview met dezelfde mensen.

Focusgroepen dienen goed voorbereid te worden. Zo is het verstandig van tevoren na te denken welke onderwerpen behandeld moeten worden. Ook de selectie van deelnemers behoeft enige voorbereiding. Voor een focusgroep is het van belang dat de deelnemers enige kennis hebben over het onderwerp. Afhankelijk van de specifieke vraag kan er gekozen worden voor een focusgroep onder respondenten met dezelfde dan wel verschillende achtergrond. In een focusgroep kan het handig zijn gebruik te maken van een externe gespreksleider. Zo heb je alle aandacht om de discussie te observeren en vast te leggen.

Relevante literatuur: Robson (2002) en Swanborn (1999).

Voorbeelden van gebruik in onderzoek

- In het onderzoek naar [bekostiging van de curatieve geestelijke gezondheid](#) is na de dataverzameling een bijeenkomst georganiseerd met sleutelfiguren uit het veld.
- Bij het onderzoek naar [passend onderwijs](#) is gebruikgemaakt van webfocusgroepen.
- Bij het onderzoek naar [CE-markering](#) zijn drie expertmeetings georganiseerd, waarin de bevindingen zijn besproken met publieke en private betrokkenen.

9.2.1.3 Enquête

9.2.1.3.1 Algemeen

Met een enquête kunnen op relatief eenvoudige wijze de meningen, attitudes of kennis van een representatieve groep respondenten over een bepaald onderwerp worden verzameld. Daarnaast kunnen in enquêtes kenmerken of gedragingen bevestigd worden die niet kunnen worden afgeleid uit registratiebestanden, maar die wel voor de uitkomsten (doelmatigheid en doeltreffendheid) van het beleid van belang kunnen zijn. De kennis kan van belang zijn als de betreffende informatie niet (voldoende) via andere bronnen beschikbaar is.

Enquêtes kunnen bestaan uit open en/of gesloten vragen. In kwantitatief onderzoek is het gebruikelijk voornamelijk gesloten vragen op te nemen omdat de resultaten dan makkelijker analyseerbaar zijn met statistische methoden.

Voor het meten van houdingen of gedrag van mensen, is de Likertschaal een zeer geschikte methode. Een Likertschaal meet houding en gedrag met behulp van antwoordkeuzen die variëren van het ene uiterste tot het andere (bijvoorbeeld van helemaal niet waarschijnlijk tot zeer waarschijnlijk). Die antwoordkeuzes worden omgezet in cijfers, waardoor de antwoorden geanalyseerd kunnen worden.

Een paar algemene aandachtspunten bij het opstellen van enquêtes zijn:

- Probeer het wiel niet opnieuw uit te vinden maar maak gebruik van bestaande inzichten. Er bestaan al heel veel enquêtes over uiteenlopende onderwerpen en er zijn diverse meetinstrumenten ontwikkeld en gevalideerd. Deze kunnen heel bruikbaar zijn. Het is aan te bevelen om deze kennis te benutten, dus raadpleeg wetenschappelijke artikelen of rapporten van onder meer rijkskennisinstellingen en zoek daar naar goede voorbeelden. Let daarbij wel goed op de kwaliteit van het onderliggende onderzoek (zie ook Normen voor de kwaliteit van derdeninformatie voor gebruik in eigen onderzoek).
- Zorg dat je per vraag één onderwerp behandelt, dus niet twee vragen in één.
- Houd de vragen kort en bondig.
- Vermijd dubbele ontkenningen.
- Leg moeilijke begrippen uit.
- Vermijd woorden die een bepaalde emotionele waarde in zich hebben.
- Vermijd suggestieve vragen (zoals: vindt u ook niet dat...?).
- Zorg voor een balans tussen positief en negatief geformuleerde vragen.
- Maak de vragenlijst niet te lang (ongeveer 10 tot 15 minuten tijd voor het invullen).
- Meld altijd, indien van toepassing, dat de vragenlijst anoniem is.
- Let op bepalingen in de AVG: als je de enquête bijvoorbeeld wil koppelen aan andere databestanden, dan moet je hiervoor in de enquête toestemming vragen van de ondervraagde.

Het is aan te raden om een enquête eerst te testen, bijvoorbeeld onder een aantal collega's en vervolgens bij de uitvoering onder een kleine groep respondenten. Dit laatste is alleen nodig als de enquête wordt uitgezet naar een grote groep respondenten. Bekijk bij zo'n test ook altijd het databestand en ga na of je met de proefdata inderdaad de onderzoeksvragen kunt beantwoorden. Idealiter heb je de analyseaanpak al klaar voordat de enquête wordt uitgestuurd.

Voorbeelden van gebruik in onderzoek

- In het onderzoek naar [stimulering van duurzame energieproductie](#) is een grootschalige enquête onder alle aanvragers van een SDE+-subsidie uitgevoerd.
- In het focusonderzoek naar [digitaal thuiswerken](#) is een online-enquête gebruikt waaraan (vrijwel) alle rijksambtenaren konden deelnemen.

9.2.1.3.2 Vignettenmethode

Bij een enquête kun je gebruik maken van de *vignettenmethode* als je inzicht wil krijgen waarom een bepaalde groep mensen bepaalde keuzes maakt. Bijvoorbeeld welke factoren het belangrijkste zijn bij de keuze van een kandidaat in een sollicitatie of welke factoren van invloed zijn op de bereidheid van mensen om mantelzorg te verlenen. Het idee achter de vignettenmethode is dat deze aanpak beter werkt dan mensen direct vragen naar het belang dat ze hechten aan bepaalde factoren, onder meer omdat het risico bestaat dat ze dan sociaal wenselijke antwoorden geven. De vignettenmethode wordt in de internationale wetenschappelijke literatuur ook wel aangeduid als *conjoint analysis*. De aanpak wordt vooral toegepast in marktonderzoek.

In de vignettenmethode worden verschillende (meestal hypothetische) situaties (vignetten) voorgelegd aan de geënquêteerde. De vignetten bestaan uit een beschrijving van een situatie op basis van de factoren die worden geacht van invloed te zijn op, in dit geval de keuze voor een bepaalde sollicitant of de mate van bereidheid om mantelzorg te verlenen. De geënquêteerde wordt vervolgens gevraagd om de situaties te scoren op een bepaalde uitkomstmaat. De constructie van de situaties is zodanig dat het mogelijk is te bepalen in welke mate de verschillende factoren van invloed zijn op de uitkomst.

Bij het gebruik van de vignettenmethode is een aantal zaken erg belangrijk. Ten eerste moet je goed nadenken over de factoren die van invloed zijn op de uitkomst. De methode kan immers alleen het effect meten van de factoren die worden meegenomen. Ten tweede moet je nadenken over de (omvang en samenstelling) van de groep die gevraagd wordt om de vignetten te beoordelen. Persoonskenmerken van de geënquêteerden kunnen namelijk van invloed zijn op de uitkomst. Om hiervoor te kunnen corrigeren, moet de groep geënquêteerden groot genoeg zijn.

Je kunt meer lezen over de vignettenmethode in De Graaf, Kok en Berkhout (2004) en Veenma, Batenburg & Breedveld (2004). Een meer technische beschrijving is te vinden in Van Praag en Frijters (1997), onder het hoofdstuk met de titel *Conjoint measurement approach* en in Hauser en Rao (2002).

9.2.1.3.3 Steekproefselectie

Vaak wordt bij het selecteren van respondenten voor enquêtes gebruikgemaakt van steekproeven. De steekproef vormt dan een weergave of representatie van de gehele populatie. Om representatief te zijn, moet de steekproef groot genoeg zijn en aan nauwkeurige statistische regels voldoen. Er is geen eenvoudige regel voor het bepalen van de steekproefomvang die voor alle enquêtes kan worden gebruikt. Zo is de steekproefomvang afhankelijk van de vraag waarvoor de resultaten van de enquête worden gebruikt. Over het algemeen leiden grotere steekproeven tot kleinere foutmarges.

Een belangrijke overweging voor de steekproefomvang is ook de noodzaak om iets te zeggen over resultaten voor subgroepen (bijvoorbeeld mannen en vrouwen, werkenden en werklozen, verschillende doelgroepen). Vervolgens moet worden gezorgd dat de steekproefomvang in elke subgroep voldoende is om geldige cijfers uit de resultaten te halen. Let dus op met resultaten over subgroepen die in je steekproef heel klein zijn, daar zijn eigenlijk geen uitspraken over te doen.

Het trekken van een steekproef uit een populatie kan op verschillende manieren:

1. **Eenvoudige willekeurige selectie:** elke eenheid wordt volledig willekeurig (random) gekozen. Dit betekent dat iedereen een even grote kans heeft om tot de steekproef te behoren. Elke eenheid kan maar 1 keer gekozen worden.
2. **Systematische selectie:** elke bijvoorbeeld tiende of honderdste eenheid wordt gekozen. De eerste eenheid wordt willekeurig gekozen.
3. **Gestratificeerde willekeurige selectie:** als de populatie niet voldoende homogeen is, wordt deze in groepen of clusters verdeeld die onderling wel (zoveel mogelijk) homogeen zijn. Vervolgens wordt uit elke groep een willekeurige selectie gemaakt. Deze aanpak wordt ook wel gebruikt als bepaalde voor het onderzoek belangrijke groepen ondervertegenwoordigd zijn in de populatie of als er verwacht wordt dat een bepaalde groep een lagere respons dan gemiddeld zal hebben. Uit deze groep wordt dan een relatief grotere groep respondenten gekozen (*oversampling*).
4. **Selectie per cluster:** als de populatie niet voldoende homogeen is, wordt de populatie in groepen of clusters verdeeld die onderling wel (zoveel mogelijk) homogeen zijn. Vervolgens wordt een willekeurige selectie van clusters getrokken en wordt het hele cluster benaderd. Clusters kunnen hier bijvoorbeeld landen zijn in een internationaal onderzoek naar meerdere landen. In het onderzoek worden

dan alle eenheden in een bepaald land benaderd om een uitspraak op het niveau van een land te kunnen doen.

5. **Selectie per quota:** de onderzoekers maken een kleinere weergave of representatie op basis van de kenmerken van de populatie. Ze moeten vervolgens eenheden of individuen vinden die overeenkomen met deze kenmerken, zodat de steekproef lijkt op de populatie. Een voorbeeld hiervan is zoeken naar de meeste 'gemiddelde' gemeente in Nederland en op basis van de situatie, stemming en/of omstandigheden in die gemeente uitspraken doen over heel Nederland.
6. **Sneeuwbalsteekproef:** in dit geval wordt aan respondenten gevraagd of zij andere personen kennen die tot de populatie behoren die wordt onderzocht. Deze manier van steekproeftrekking wordt vaak toegepast bij groepen die bij een andere vorm van steekproeftrekking moeilijk vindbaar of benaderbaar zijn. Er zitten wel nadelen aan deze methode: het succes is erg afhankelijk van de grootte van het netwerk van de respondenten die als eerste worden benaderd. Mensen met een groot netwerk hebben een grotere kans om in de steekproef terecht te komen en er bestaat een risico dat je steekproef alleen bestaat uit bepaalde clusters in de populatie.

Als je op basis van de steekproef uitspraken wil doen over de hele populatie, verdient het aanbeveling te werken met een eenvoudige willekeurige selectie. Alle andere selectiemethoden vereisen specifieke berekeningen om de uitkomsten te kunnen veralgemeniseren of zijn ongeschikt om een generaliserende uitspraak te doen. Bij de steekproeftrekking kan gebruikgemaakt worden van software, zoals R of SPSS.

9.2.2 Bewijs uit documenten

Bij het beantwoorden van veel onderzoeksvragen speelt bewijs uit documenten een belangrijke rol. Er kan onderscheid worden gemaakt tussen de volgende soorten documenten:

- (openbare) beleidsstukken en wetgeving;
- jaarverslagen;
- interne notities, memo's, verslagen, etcetera;
- bestaande statistieken;
- archiefstukken;
- externe evaluaties, onderzoeken en (wetenschappelijke) publicaties;
- bestaande databases (zowel interne databases als open data). Mocht je verschillende databases of databronnen aan elkaar willen koppelen, let dan goed op de definities van bepaalde begrippen.
- Let bij het opvragen van gegevens op de relatie met de onderzoeksvraag.

De informatie uit documenten kan bijvoorbeeld gebruikt worden om een reconstructie te maken van de beleidstheorie (9.2.2.1) en kan de basis zijn voor een literatuuronderzoek of meta-onderzoek (9.2.2.2).

9.2.2.1 Reconstructie van de beleidstheorie

Kennis van de beleidstheorie is van belang in ieder onderzoek waarin we onszelf de vraag stellen “werkt het beleid zoals bedoeld?”. Vaak is echter niet direct helder hoe het beleid precies bedoeld is. In die gevallen kan een methode om de theorie achter het beleid te reconstrueren en eventueel te beoordelen uitkomst bieden.

Hoogerwerf (1989) definieert een beleidstheorie als “het geheel aan veronderstellingen van een actor dat aan een beleid ten grondslag ligt”. Deze definitie geeft aan dat een beleidstheorie zowel gebonden is aan een bepaald beleid als aan een bepaalde actor. Deze ‘actor’ kan een persoon zijn, maar ook een organisatie. Zo kan gesproken worden van de beleidstheorie van een ambtenaar of een subsidieverstrekker, maar ook van de beleidstheorie van het Ministerie van Economische Zaken voor bepaald beleid.

Als Algemene Rekenkamer gaan we over het algemeen uit van de officiële theorie van een ministerie, respectievelijk de minister. In veel gevallen is deze theorie echter niet of heel impliciet aanwezig en moet deze eerst gereconstrueerd worden.

Wanneer de theorie helder is, kan deze ook beoordeeld worden op zaken als empirisch gehalte of consistentie. Dit kan informatie opleveren over de sterktes en zwaktes van het beleid.

Zie ook **criteria** voor het beoordelen van een beleidstheorie.

Voorbeeld van gebruik in onderzoek

In het onderzoek naar het [energielabel voor koopwoningen](#) is de gedragstheorie gereconstrueerd en beoordeeld..

9.2.2.2 Literatuuronderzoek en meta-onderzoek

Literatuuronderzoek en meta-onderzoek kunnen worden gebruikt om een specifieke onderzoeksvraag te beantwoorden. Literatuuronderzoek is ook van belang in de opstartfase van een onderzoek om bijvoorbeeld meer inzicht te verkrijgen in een onderwerp.

Literatuuronderzoek en meta-onderzoek zijn erop gericht op een systematische manier de stand van zaken op te maken van een bepaald onderwerp. Hiervoor kan bijvoorbeeld worden gebruikgemaakt van wetenschappelijke artikelen, krantenartikelen, websites, boeken, Kamerstukken en wetteksten.

Meta-onderzoek is een specifieke invulling van literatuuronderzoek. In meta-onderzoek worden onderzoeken van een bepaald fenomeen samengevoegd of

vergeleken om één secuurdere uitkomst te verkrijgen. Door de resultaten uit eerdere onderzoeken gezamenlijk te bekijken kunnen uitspraken worden gedaan en inzichten worden verkregen die op basis van elk van de afzonderlijke onderzoeken niet mogelijk waren.

Meta-onderzoek bestaat uit meta-evaluatie en evaluatiesynthese. Meta-evaluatie is een evaluatie van al bestaande evaluaties en geeft antwoord op de vragen: welke verschillen zijn er in de uitkomsten van de evaluaties en hoe zijn deze verschillen te verklaren? Evaluatiesynthese daarentegen probeert de uitkomsten van verschillende onderzoeken te aggregeren (samen te vatten) en aan de hand daarvan een uitspraak te doen over een bepaald onderwerp. Bij evaluatiesynthese vindt vaak een statistische heranalyse plaats, soms op basis van de brondata en soms op basis van al geaggregeerde gegevens zoals deze in wetenschappelijke publicaties over de oorspronkelijke studies verschenen zijn. Statistische heranalyse is geen vereiste bij meta-evaluaties, want ook op kwalitatieve gegevens kan een meta-evaluatie plaatsvinden.

Twee interessante bronnen met veel voorbeelden van meta-onderzoek zijn de [Campbell Collaboration](#), gericht op sociale wetenschappen en het [Cochrane Network](#), gericht op de gezondheidszorg.

Voorbeelden van onderzoek: een meta-onderzoek naar de kwaliteit en inzichten uit wetsevaluaties (Klein Haarhuis en Niemijer 2008), een meta-onderzoek naar de effecten van preventie van geweld in de (semi)publieke ruimte (Van der Knaap en Bogaerts, 2006) en een meta-analyse van Nederlands recidiveonderzoek naar de effecten van strafrechtelijke interventies (Wartna, Alberda en Verweij, 2013).

9.2.3 Fysiek bewijs

Het kan in sommige gevallen zinvol zijn om bewijs uit interviews of bewijs uit documenten te verifiëren aan de hand van fysiek bewijs. Andersom kan fysiek bewijs ook worden gebruikt in interviews of om beschrijvingen in documenten te bevestigen of te weerleggen. We onderscheiden observatie en inspectie van fysieke objecten:

- **Observatie.** Hierbij wordt de situatie in de praktijk geobserveerd door de onderzoekers. Een lichte vorm van observatie is een werkbezoek, een meer uitgebreide vorm is een langduriger observatie van de manier van werken (participatieve observatie). Een bijzondere vorm van participatieve observatie is toegepast in het onderzoek naar [digitale dijkverzwinging](#). In dit onderzoek is de cybersecurity van vitale waterwerken in de praktijk getoetst met hulp van ethische hackers.
- **Inspectie van fysieke objecten.** Hierbij wordt meestal een beperkt aantal objecten bekeken, om hier gedetailleerde informatie over te verzamelen. Deze aanpak is onder meer toegepast in de beleids casus [Gebruiksbeperkingen van hoofdvaarwegen](#) in het Verantwoordingsonderzoek 2018.

9.3 Methoden en technieken van data-analyse

In een data-analyse worden de verzamelde data gecombineerd, bewerkt en onderzocht. Zo kan daaruit de informatie worden gedestilleerd die nodig is om antwoord te geven op de onderzoeksvragen.

Het is belangrijk om vooraf goed na te denken over het onderzoeksontwerp. Dit ontwerp hangt nauw samen met de onderzoeksvragen. Als je in het onderzoek bijvoorbeeld op zoek bent naar verklaringen voor of oorzaken van een bepaald fenomeen, zal je onderzoeksmethoden moeten gebruiken die op gestructureerde wijze helpen om oorzaken in beeld te brengen. In de meeste gevallen is het opstellen van een onderzoeksontwerp bij het onderzoeken van doeltreffendheid en doelmatigheid maatwerk, omdat deze niet alleen zal afhangen van de specifieke onderzoeksvragen, maar ook van de specifieke kenmerken van het beleidsterrein en de beschikbaarheid van gegevens.

Hier zetten we een aantal analysemethoden voor doeltreffendheids- en doelmatigheidsonderzoek op een rij. Allereerst bespreken we een aantal standaard analysemethoden die vaak worden gebruikt in rekenkameronderzoek (9.3.1).

Vervolgens gaan we in op een aantal analysemethoden voor doeltreffendheid en doelmatigheid die zijn toegepast of die in potentie in rekenkameronderzoek zouden kunnen worden toegepast. Hierbij kijken we allereerst naar analysemethoden waarmee we de doeltreffendheid van beleid kunnen onderzoeken (9.3.2). Daarna richten we ons op analysemethoden waarmee we de doelmatigheid van beleid kunnen onderzoeken (9.3.3). Afsluitend bieden we een overzicht van een aantal interessante internetbronnen waarin je meer kan lezen over het gebruik van onderzoeksmethoden en een overzicht van relevante literatuur (9.3.4).

Vooraf moet worden opgemerkt dat de gepresenteerde methoden geen uitputtende lijst van mogelijkheden bieden. Ook worden de verschillende methoden slechts op hoofdlijnen besproken. Het overzicht hier is dus vooral bedoeld als inspiratie. Bij alle analysemethoden geven we zo veel mogelijk voorbeelden uit de dagelijkse praktijk van rekenkamer- of andersoortig onderzoek. Op diverse plaatsen verwijzen we naar achterliggende literatuur die meer informatie geeft over de desbetreffende methoden.

9.3.1 Standaard analysemethoden in rekenkameronderzoek

In rekenkameronderzoeken maken we in veel gevallen gebruik van beschrijvende data-analyse en *process tracing*. Beschrijvende data-analyse (9.3.1.1) is met name toepasbaar in onderzoek waarin je een bepaalde situatie wil beschrijven. *Process tracing* (9.3.1.2) is met name bruikbaar als je op zoek bent naar verklaringen voor de aanwezigheid van bepaalde uitkomsten, maar kan ook worden aangewend om bepaalde (beleids)processen te beschrijven.

9.3.1.1 Beschrijvende data-analyse

Vaak heeft onderzoek het doel om inzicht te geven in aspecten die belangrijk zijn om (de gevolgen van) een bepaald beleid of beleidswijziging te begrijpen. Hoewel bestaande publicaties, zoals jaarverslagen van ministeries of uitvoerders, veel informatie bevatten, geven deze vaak niet voldoende inzicht voor het beantwoorden van onze onderzoeksvragen. Op zo'n moment heeft het gebruik van beschrijvende data-analyse meerwaarde.

Bij de toepassing van beschrijvende data-analyse kiezen we ervoor om voor een groot aantal eenheden een beperkt aantal kenmerken te onderzoeken. Hierbij gaat het erom de empirische werkelijkheid beter inzichtelijk te maken aan de hand van een aantal analytische dimensies die gerelateerd zijn aan deze kenmerken. Hierbij kunnen we eenheden aan de hand van één (univariaat) of meerdere (multivariaat) kenmerken beschrijven en kunnen we op basis van deze kenmerken bijvoorbeeld tijdreeksen, classificaties of rankings opstellen. De verdeling van kenmerken langs dimensies geeft ons inzicht in patronen. Deze patronen kunnen instrumenteel zijn om beschrijvende uitspraken te doen over de praktijk. De praktijk kunnen we in sommige gevallen ook toetsen aan een norm. Door toepassing van dit type analyse kunnen we dus zowel antwoorden formuleren op beschrijvende als toetsende onderzoeksvragen.

In ons onderzoek maken we veel gebruik van datavisualisaties als specifieke invulling van beschrijvende data-analyse. Hierbij is het belangrijk om het juiste type figuur te kiezen voor het inzicht dat je wilt bieden. Een mooi overzicht kun je vinden in de [R-graph Gallery](#). Hier worden onder andere voorbeelden gegeven van figuren die geschikt zijn om verdeling (*distribution*), onderverdelingen (*part of a whole*), verschillen (*ranking*), ontwikkeling over tijd (*evolution*) of geografische verschillen (*map*) in beeld te brengen. Visualisaties kunnen in rapporten worden opgenomen, maar ook interactief in een dashboard zoals we met de coronarekening hebben gedaan.

Het gebruik van beschrijvende data-analyse kent ook beperkingen: zo kan je op basis van beschrijvende data-analyse geen uitspraken doen over oorzaken van bepaalde uitkomsten. Wel kunnen resultaten uit een beschrijvende analyse een bouwsteen zijn voor verklarende uitspraken. Hiervoor dient deze methode bijvoorbeeld te worden gekoppeld aan de toepassing van *process tracing* of aan meer geavanceerde analysemethoden. Je kunt op basis van beschrijvende data-analyse geen uitspraken doen over de toegevoegde waarde van beleid, omdat je niet hebt gecorrigeerd voor het effect van andere factoren.

Voorbeelden van beschrijvende data-analyse in een beschrijvend onderzoek

In het onderzoek naar [toeslagen terugbetalen](#) is per huishouden gekeken naar de hoogte, de duur, de stapeling en de opeenvolging van toeslagschulden. Het departement had daarvoor alleen gerapporteerd over individuele toeslagen voor individuele toeslagontvangers. Door de data op het niveau van huishoudens te bekijken en de samenloop van verschillende toeslagen bij een huishouden te beschouwen, waren we in dat onderzoek in staat te laten zien dat het probleem van toeslagschulden in de tijd stapelt bij bepaalde groepen huishoudens.

In het onderzoek naar [sociale zekerheid en flexibele arbeidsmarkt](#) is onderzocht in welke mate werkenden een tijdelijk inkomensverlies kunnen opvangen. Hierbij is onderscheid gemaakt tussen werkenden met een flexibel arbeidscontract, zelfstandigen en mensen met een vast contract. Er bleken aanzienlijke verschillen te bestaan, waarbij dit vermogen bij werkenden met een flexibel arbeidscontract relatief laag bleek te zijn.

Voorbeeld van beschrijvende data-analyse in een toetsend onderzoek

Als onderdeel van een onderzoek naar het [UWV](#) (*UWV, balanceren tussen ambities en middelen*) is beoordeeld of het UWV in staat is geweest om zijn taken binnen de afgesproken tijd met beschikbare mensen en middelen te realiseren. Om te bepalen of het UWV hiertoe in staat is geweest, wordt onder andere getoond dat herbeoordelingen van arbeidsongeschikten gedurende de onderzoeksperiode in afnemende mate tijdig zijn uitgevoerd (87% van de gevallen in 2015). Deze bevinding wordt, samen met andere informatie, gebruikt om een toetsende onderzoeksvraag te beantwoorden, namelijk of het UWV erin is geslaagd zijn taken binnen de afgesproken tijd te realiseren.

9.3.1.2 Process tracing

Als we verklarende onderzoeksvragen beantwoorden, gaan we op zoek naar hoe een oorzaak met een effect in verband kan worden gebracht. Een vaak gehanteerde analysemethode is in dit geval *process tracing*. De kern van *process tracing* is om stap voor stap op zoek te gaan naar informatie die van het effect leidt naar een mogelijke oorzaak voor dat effect. Om dit te bereiken breng je aan de hand van data tussenstapjes (ook wel interveniërende variabelen genoemd) in kaart. Door de tussenstappen in kaart te brengen, krijg je inzicht in de causale mechanismen die verantwoordelijk zijn voor bepaalde uitkomsten. Op deze manier is *process tracing* een zeer geschikte methode om verklaringen bloot te leggen voor de gerealiseerde mate van doeltreffendheid of doelmatigheid.

Bij de toepassing van *process tracing* maken we onder andere gebruik van gedachte-experimenten aan de hand van counterfactuals: wat zou er gebeuren als X1 niet zou hebben plaatsgevonden? Zou X2 dan ook niet hebben plaatsgevonden? En daarmee uiteindelijk Y ook niet tot stand zijn gebracht? Als het antwoord op deze vragen bevestigend is, betekent dit dat X1 waarschijnlijk een onderdeel is van het aanwezige causale mechanisme. Een ander hulpmiddel bij process tracing is het kijken naar neveneffecten. Hierbij kan je voor een vermoedelijke oorzaak nagaan wat voor type specifieke neveneffecten je in zo'n geval zou verwachten. Het vinden van dergelijke effecten ondersteunt je verwachting dat deze oorzaak verantwoordelijk is voor een uitkomst. In het verlengde hiervan kan je ook kijken naar het uitblijven van neven-effecten als er alternatieve oorzaken aan de orde zouden kunnen zijn. Als je een dergelijk neveneffect niet vindt, helpt je dit om de alternatieve oorzaak te verwerpen.

Voorbeelden

Het onderzoek naar [CE-markering](#) laat zien hoe het komt dat producten die niet voldoen aan de Europese CE-regelgeving toch op de Europese markt komen en wat de overheid doet om dit te voorkomen en te verhelpen. Een deel van het onderzoek bestond uit een productreis voor twee producten, waarbij de hele keten (van producent tot consument) in kaart is gebracht. Bij elke stap in de keten is gekeken wat de betreffende partij heeft gedaan voor controle op naleving van de Europese CE-regelgeving. Dit is gedaan door interviews te houden en documenten op te vragen van producent tot eindgebruiker.

In het onderzoek [Uit het vizier](#) is aan de hand van vier casussen gekeken naar de besluitvorming over en uitvoering van de bezuinigingsoperatie bij het ministerie van Defensie in het afgelopen decennium. Door voor elk van deze vier casussen het gehele proces van aanloop tot het bezuinigingsbesluit tot de feitelijke implementatie te volgen, wordt geconcludeerd dat in geen van de vier gevallen de bezuinigingen daadwerkelijk werden gerealiseerd. In een aantal gevallen was er sprake van het terugdraaien van bezuinigingen zonder dat de Tweede Kamer daar adequaat over was geïnformeerd.

Een andere toepassing van *process tracing* is *process verification*. In de context van rekenkameronderzoek toetsen we de beleidstheorie vaak aan de praktijk. Hierbij expliciteren we eerst de verschillende stappen die de inzet van een instrument koppelen aan het behalen van een beleidsdoel (verondersteld causaal mechanisme). Vervolgens toetsen we de beleidstheorie om zo te achterhalen of het veronderstelde causale mechanisme ook is terug te vinden in de praktijk. Door toepassing van *process verification* kan je de doeltreffendheid van de inzet van een bepaald instrument voor het behalen van een beleidsdoel aannemelijk maken. Als onderdeel

van *process verification* probeer je ook alternatieve oorzaken voor het bereiken van een bepaald effect uit te sluiten. Dit doe je door bewijs te vinden voor de afwezigheid van een verband tussen deze alternatieve oorzaken en het effect.

Een sterk aan *process tracing* verwante kwalitatieve analysemethode is incidentanalyse. Hierbij ga je op zoek naar oorzaken die hebben geleid tot een ongewenste uitkomst. Een veelgebruikt instrument om een incidentenanalyse uit te voeren is *Tripod Bèta*. De Tripod Bèta-methode stimuleert onderzoekers om te zoeken naar eventuele latente fouten in een organisatie, die hebben bijgedragen aan het ontstaan van een incident of een verstoring. Het hoofddoel van een incidentanalyse is herhaling voorkomen. Dat doe je door tekortkomingen te identificeren en aan te geven welke maatregelen moeten worden genomen om het incident of de verstoring in de toekomst te voorkomen. Incidenten of verstoringen worden veroorzaakt door een reeks gebeurtenissen. Zulke reeksen kunnen in verschillende fasen van het ontstaan van het incident geïdentificeerd worden. Het onderzoek dient zich niet te beperken tot de actieve fouten, aangezien dit vaak niet de feitelijke oorzaak is. De Tripod Bèta-methode gaat ervan uit dat onjuiste handelingen niet op zichzelf staan, maar worden beïnvloed door omgevingsfactoren, die het gevolg kunnen zijn van fouten in de organisatie. Dit zijn verborgen gebreken. Dit type gebreken (latente fouten) kunnen lange tijd verborgen blijven binnen het systeem. Uit incidentenonderzoek blijkt vaak dat incidenten niet voorkomen worden door het corrigeren van actieve fouten maar door het sturen op latente fouten. Deze latente fouten kunnen onder andere betrekking hebben op procedures, training, communicatie en tegenstrijdige doelstellingen.

De latente fouten en de context waarin deze plaatsvinden worden in rekenkamertermen geïnterpreteerd als de verklaring voor de achtergebleven prestatie van de onderzochte organisatie.

Voorbeelden

In een onderzoek naar het [UWV](#) (*UWV, balanceren tussen ambities en middelen*) is gekeken naar de verbetermaatregelen die zijn genomen om de achterstanden bij de herbeoordelingen volgens de Wet werk en inkomen naar arbeidsvermogen (WIA) weg te werken. In het onderzoek is nagegaan in hoeverre deze verbetermaatregelen de achterliggende oorzaken voor het ontstaan van deze achterstanden wegnemen.

In het onderzoek naar [bescherming van natuurgebieden](#) was het incident het niet werken volgens de wet- en regelgeving. Aan de hand van een aantal geselecteerde casussen is onderzocht in hoeverre bij aantasting van natuur door aanleg van woningen, infrastructuur, bedrijventerreinen en dergelijke de verplichte stappen in het kader van natuurbeschermingsregels ook daadwerkelijk door de betrokken actoren zijn gezet.

Process tracing kan ook worden gebruikt in beschrijvend onderzoek. In zo'n geval breng je voor een bepaald (beleids)proces de verschillende tussenstappen in kaart om op deze manier het proces inzichtelijk te maken voor jezelf en je publiek.

Visuals kunnen hierbij bijdragen aan een toegankelijke presentatie.

Bij het toepassen van *process tracing* moet je, nog meer dan bij het hanteren van kwantitatieve methoden, gedurende de uitvoering van het onderzoek meerdere malen opnieuw data verzamelen, analyseren en reflecteren en vele malen van de ene naar de andere fase heen en weer bewegen. Onderzoek op basis van *process tracing* heeft dus een sterk iteratief karakter. Een belangrijke beperking van *process tracing* is dat je meestal maar een beperkt aantal casussen kunt onderzoeken; dit kan een belangrijke beperking zijn voor de generaliseerbaarheid van de uitkomsten.

9.3.2 Analysemethoden voor onderzoek naar doeltreffendheid

Voor zowel beschrijvende data-analyse als *process tracing* geldt dat het tot zeer waardevolle bevindingen kan leiden. Maar om robuustere uitspraken te kunnen doen over doeltreffendheid is het vaak nodig om meer geavanceerde analysemethoden toe te passen, met name als er geen evident bewijs is dat het beleid geen resultaten oplevert. Hier wordt een aantal analysemethoden besproken, waarbij we grofweg de volgorde van minst naar meest complex hanteren.

9.3.2.1 Qualitative Comparative Analysis (QCA)

Bij gebruik van Qualitative Comparative Analysis (QCA), in het Nederlands ook bekend als systematisch kwalitatieve vergelijkende analyse, ga je op zoek naar combinaties van oorzaken ('condities' in QCA-termen) die leiden tot een bepaald effect ('uitkomst' in QCA-termen). Hierbij ga je na welke (combinatie van) condities een voldoende of noodzakelijke voorwaarde is voor de aan- of afwezigheid van een uitkomst. Toepassing van QCA is vooral interessant als je over een redelijk aantal casussen beschikt (zo tussen 10 en 50). Dit aantal casussen is vaak te groot om uitgebreid kwalitatief diepteonderzoek te doen en te klein om kwantitatieve analysemethoden toe te passen.

QCA heeft een specifieke opvatting over causaliteit die gebaseerd is op *booleaanse algebra*. Deze opvatting wordt er allereerst door gekenmerkt dat oorzaken vaak samengesteld zijn; het is de specifieke combinatie van condities die uiteindelijk leiden tot een uitkomst (causaliteit is 'conjunctureel'). Daarnaast kunnen de oorzaken voor de aanwezigheid en de afwezigheid van een uitkomst van elkaar verschillen; oorzaken voor de afwezigheid van een uitkomst zijn niet noodzakelijk het spiegelbeeld van de oorzaken van de aanwezigheid van een uitkomst (causaliteit is asymmetrisch). Afsluitend zijn er verschillende combinaties van condities mogelijk die tot een bepaalde uitkomst leiden (equifinaliteit).

QCA kan worden ingezet om de doeltreffendheid van beleid te onderzoeken. Deze benadering sluit aan bij het denken over evaluaties vanuit zogenoemd 'realistisch' perspectief, waarin nadrukkelijk rekening wordt gehouden met de interactie tussen instrumenten en de context van beleid en uitvoering. Het beleidsdoel (uitkomst) wordt daarbij behaald door de inzet van verschillende instrumenten in een bepaalde mix of context (combinaties van condities). Bij toepassing van varianten van QCA wordt vaak gebruikgemaakt van analyseprogramma's als *Tosmana* en R. QCA kent echter ook een aantal belangrijke beperkingen. Allereerst is het aantal verklarende condities dat je kan meenemen in je analyse beperkt. Als je dus nog geen helder beeld hebt van wat de mogelijke oorzaken van je effect zijn (bijvoorbeeld contextfactoren die invloed hebben op de werkzaamheid van je beleidsinstrument), is het beter om eerst een diepgaande casestudy uit te voeren. Daarnaast noopt QCA tot het terugbrengen van de 'informatierijkheid' van je data. In de standaardtoepassing van QCA, de zogenoemde *crisp set* QCA, dien je de scores op je condities en uitkomst te classificeren als aan- of afwezig. Hierdoor kan er waardevolle informatie over je casussen verloren gaan.

Er bestaat overigens wel een meer geavanceerde vorm van QCA, *fuzzy set* QCA, waarin dit probleem wordt ondervangen en waarden niet hoeven te worden geclassificeerd in een aan- of afwezigheidsgroep. Dit gaat echter vaak wel gepaard met moeilijker te interpreteren analyse-uitkomsten.

Voorbeeld

Het onderzoek naar [inburgering](#) richt zich op de vraag in hoeverre het inburgeringsbeleid zoals dat in 2013 is ingevoerd ertoe leidt dat migranten die onder de wet vallen daadwerkelijk inburgeren en wat dit het Rijk kost. Als onderdeel van dit onderzoek zijn 38 inburgeraars geïnterviewd. De data afkomstig uit deze semigestructureerde interviews zijn vervolgens gebruikt om door toepassing van QCA te achterhalen welke (combinatie van) condities noodzakelijk of voldoende zijn voor de inburgeraar voor het wel of niet slagen voor het inburgeringsexamen (uitkomst). Uit deze analyse kwam onder andere naar voren dat een drietal configuraties leidde tot het slagen voor het inburgeringsexamen: 1) gemotiveerd zijn, een snelle start en goed welzijn, 2) gemotiveerd zijn, contact met Nederlanders en goed welzijn en 3) gemotiveerd zijn, hoge opleiding, contact met Nederlanders en snelle start. Daarnaast werden er ook verschillende configuraties gevonden (10 in totaal) die leidden tot het niet slagen voor het inburgeringsexamen.

Literatuur: Verweij & Gerrits (2012). Systematisch kwalitatief vergelijkende analyse. KWALON – Tijdschrift voor Kwalitatief Onderzoek, 17(3), 25-33.

9.3.2.2 Regressieanalyse

In een regressieanalyse wordt de samenhang bekeken tussen enerzijds een afhankelijke variabele (bijvoorbeeld het inkomen van een persoon) en een aantal onafhankelijke variabelen (bijvoorbeeld opleiding en aantal jaren werkervaring). Een regressie is een krachtig instrument om deze samenhang te onderzoeken. Als er wordt gekeken naar één afhankelijke variabele en één onafhankelijke variabele, spreken we van een enkelvoudige regressie. Meer gebruikelijk is het om te werken met één afhankelijke variabele en meerdere onafhankelijke variabelen. In dat geval spreken we over een *meervoudige regressieanalyse*.

Een belangrijke beperking van een regressieanalyse is dat er alleen een uitspraak gedaan kan worden over samenhang (correlatie), niet over causaliteit. Als je in een regressieanalyse een sterke samenhang vindt tussen bijvoorbeeld opleiding en inkomen, kun je niet op basis van alleen deze regressieanalyse concluderen dat er een oorzakelijk verband is. Hiervoor heb je onder meer een stevige theoretische basis en aanvullend bewijs nodig.

Voorbeeld

In het onderzoek Toegang tot langdurige zorg is onder andere gekeken naar het verband tussen aanmelder (patiënt zelf versus zorgprofessional) en de kans op toelating tot financiering van langdurige zorg. Bij het uitvoeren van deze analyses was het van groot belang om te corrigeren voor de leeftijd van de persoon voor wie financiering werd aangevraagd, omdat oudere mensen niet alleen meer zorg nodig hebben, maar ook vaker door een zorgprofessional worden aangemeld. Door een meervoudige regressie toe te passen werd voor dit soort contextuele effecten gecorrigeerd en kon de relatie tussen beleid en uitkomst worden geschat.

Als je onderzoek wil doen naar de doeltreffendheid van een concrete beleidsmaatregel die wordt ingezet om een doelstelling te bereiken, zou je dit het liefst doen aan de hand van een zuiver experimenteel design. In de context van rekenkameronderzoek is een dergelijk design echter niet haalbaar en zijn een natuurlijk experiment of een quasi-experiment de beste alternatieven. Deze aanpakken geven een betrouwbare inschatting of beleid al dan niet doeltreffend is, maar stellen wel hoge eisen aan de kwaliteit van het design (is het mogelijk om twee vergelijkbare groepen te construeren waarbij bij de ene groep de beleidsmaatregel is toegepast en bij de andere groep niet?) en aan de data (zijn er voldoende gegevens beschikbaar over de twee groepen?). Bij deze aanpak is het belangrijk om steeds te blijven reflecteren op de plausibiliteit van alternatieve (causale) verklaringen, omdat je hiervoor nooit volledig kunt corrigeren.

Regressieanalyses zijn geschikt om uitspraken over verbanden tussen variabelen (bijvoorbeeld tijd of ontvangen subsidie enerzijds en prestatie anderzijds) te onderbouwen. Met de uitkomsten van regressieanalyses kun je laten zien in welke mate variabelen aan elkaar gerelateerd zijn. Een probleem van vergelijkingen in data-analyse in niet experimentele onderzoeksontwerpen is dat er tussen de eenheden die je vergelijkt vaak meer verschillen bestaan dan alleen de beleidsinterventie. Deze verschillen kunnen ook een verklaring bieden voor een wijziging van de uitkomst. De vraag is dan in hoeverre het beleid of deze (context-)verschillen de uitkomsten verklaren.

Een manier om hiermee om te gaan, is door de uitkomsten te vergelijken in de tijd. Als je gebruikmaakt van een *herhaalde metingen regressieanalyse* (ook wel paneldata-analyse genoemd) bij het onderzoeken van doeltreffendheid, ga je eenheden met zichzelf vergelijken door ze te volgen in de tijd. Het idee is dat je op die manier het gedrag van de eenheden voor en na de beleidsinterventie met elkaar vergelijkt. De aanname hier is dat in de tijd geen andere veranderingen die van belang zijn voor een uitkomst hebben plaatsgevonden. Ook moet je over genoeg waarnemingen per eenheid beschikken.

Voorbeeld

In het onderzoek Convenanten uitgaansgeweld is gekeken wat het effect is van convenanten op het niveau van uitgaansgeweld. Deze convenanten zijn afspraken op lokaal niveau tussen publieke en private partijen om uitgaansgeweld terug te dringen. Deze convenanten stonden indertijd centraal in de aanpak van uitgaansgeweld. Ook hier worden gemeenten met zichzelf vergeleken door te kijken wat het niveau is van uitgaansgeweld voor en na het afsluiten van een convenant tegen uitgaansgeweld.

Een tweede probleem dat we vaak in beleidsonderzoek tegenkomen is dat eenheden geclusterd zijn in organisatorische of regionale groepen. *Multilevel regressie-analyse* is een specifieke vorm van regressieanalyse waarbij je rekening houdt met het feit dat (de effecten van) beleid geclusterd zijn binnen groepen of regio's. Denk bijvoorbeeld aan leerlingen op dezelfde school: de leerlingen zijn verschillend, maar ze zitten allemaal op dezelfde school. Door de clustering kan het probleem ontstaan dat er in bepaalde groepen of regio's geen onafhankelijke waarnemingen meer mogelijk zijn voor afzonderlijke eenheden. Bij het voorbeeld van de leerlingen in dezelfde klas kan er sprake zijn van een 'klaseffect', waardoor leerlingen van klas A significant anders scoren op schoolprestaties dan leerlingen die in een andere klas zitten. Een *multilevel* aanpak houdt in dat de analyse rekening houdt met deze 'gelaagdheid' in de data, door het effect van de groepen (in het voorbeeld het klaseffect) mee te nemen.

Voorbeeld

In het onderzoek [Toegang tot langdurige zorg](#) is onderzocht of de kans voor mensen om toegang te krijgen tot de Wet langdurige zorg (Wlz) verschilt per gemeente. Omdat de aanvraag voor Wlz afhankelijk is van gemeentelijk beleid was het belangrijk om rekening te houden met zowel verschillen tussen aanvragers als verschillen tussen gemeentes waar de aanvraag plaatsvond. Dit hebben we gedaan door een *multilevel* model te gebruiken waarin we rekening houden met het feit dat aanvragers binnen dezelfde gemeente een cluster vormen.

In sommige gevallen vallen de clusters samen met de eenheden die je onderzoekt. Dit is bijvoorbeeld het geval als je naar bedrijven kijkt die onderling zo verschillend zijn, dat je ze alleen met zichzelf kunt vergelijken. In die gevallen kan je een zogenoemd *fixed effects* model toepassen. In dat geval wordt er niet meer vergeleken op het hogere niveau. In het geval van bedrijven worden de bedrijven allemaal afzonderlijk bekeken en dus niet meer in samenhang.

Voorbeeld

In het onderzoek [Handhaven in het duister](#) is gekeken naar het effect van handhaving op het gedrag van bedrijven. Onder gedrag wordt verstaan of de bedrijven wel of niet een milieuovertreding begaan. Hierbij wordt onderscheid gemaakt tussen verschillende niveaus van ernst van de overtreding. Het effect van handhaving is onderzocht door bedrijven over de tijd te volgen en door middel van een regressieanalyse met zichzelf te vergelijken. Op die manier wordt gekeken of het gedrag van een bedrijf voor een handhavingsactie anders is dan daarna. In de modelspecificatie wordt rekening gehouden met de geneste structuur van de data (inspecties met overtredingen binnen bedrijven) en worden *fixed effects* geschat. Daarmee wordt het effect van handhaving per bedrijf vastgesteld.

Een andere manier om te corrigeren voor het feit dat er bij onderzoek (achteraf) van beleidseffecten vaak geen data van een goede controlegroep bestaat, is *propensity score matching*. Bij deze techniek verzamel je data van zowel mensen (of organisaties) die de beleidsmaatregel hebben ondergaan (groep A) als mensen (of organisaties) die de beleidsmaatregel niet hebben ondergaan (groep B). Het idee achter *propensity score matching* is dat je op een bepaalde manier 'koppels' probeert te vormen van iemand uit groep A en iemand uit groep B, waarbij de gekoppelden op elkaar lijken.

De kern van deze methode is het bepalen van de koppels: bij het bepalen van de koppels gaat het niet om de vraag of de koppels qua achtergrondkenmerken op elkaar lijken, maar om de vraag of ze qua kans op het ondergaan van het beleid op elkaar lijken. Bijvoorbeeld: je wil iemand uit groep A die 80% kans heeft om het beleid te ondergaan vergelijken met iemand uit groep B die 80% kans heeft om het beleid niet te ondergaan.

Het ligt voor de hand dat de mensen (of organisaties) die een grotere kans hebben om het beleid te ondergaan, oververtegenwoordigd zijn in de groep die het beleid ook daadwerkelijk heeft ondergaan. Als je hier niet voor corrigeert, zal je geen goede schatting kunnen maken van het effect van de beleidsmaatregel. Er zijn verschillende manieren waarop je de *propensity score* kunt gebruiken om deze correctie te maken. De variant van het werken met koppels biedt meestal weinig uitkomst als de kans om de interventie mee te maken bijzonder laag of hoog is, omdat er dan meestal geen *counterfactuals* te vinden zijn. Een andere toepassing van *propensity score* is om voor iedereen in groep A en B een *propensity score* te berekenen en vervolgens subgroepen van gelijke grootte te maken van individuen met een vergelijkbare *propensity score*. Dit wordt ook wel stratificatie op de *propensity score* genoemd. In veel gevallen wordt hierbij gekozen voor 5 groepen (kwintielen). Vervolgens wordt per kwintiel een vergelijking gemaakt tussen de personen in dit kwintiel die uit groep A komen en de personen in dit kwintiel die uit groep B komen.

Het berekenen van de *propensity score* vindt (meestal) plaats door een *logistische regressie* uit te voeren, waarbij groep A de 'enen' vormt (zij hebben het beleid immers ondergaan) en groep B de 'nullen'. Allerlei achtergrondkenmerken worden hierbij gebruikt als verklarende variabelen in de regressie.

Een beperking van deze methode is dat je voldoende relevante informatie moet hebben om de matching toe te kunnen passen. Als je niet alle belangrijke kenmerken meeneemt die van invloed zijn op de kans om het beleid te ondergaan, zullen je uitkomsten geen goede weergave geven van het effect van het beleid. Naast veel variabelen zijn er voor deze methode ook veel eenheden nodig, om voldoende vergelijkbare casussen uit beide groepen te hebben. In de praktijk zie je deze techniek dus toegepast worden in situaties met veel variabelen en een veel waarnemingen.

Voorbeeld: Het effect van “basisinkomen” op de bereidheid om te werken

Nobelprijswinnaar Guido Imbens heeft diverse artikelen geschreven om de techniek achter *propensity score matching* te verbeteren.

In het artikel *Matching methods in practice* (2015) geeft hij, naast een beschrijving van de techniek, drie voorbeelden van matching op basis van de *propensity score*. Het eerste voorbeeld gaat over een dataset van deelnemers aan een loterij, waarbij de winnaars 20 jaar lang jaarlijks een groot geldbedrag ontvangen. De uitkomsten van dit onderzoek zijn eerder uitgebracht in Imbens, Rubin en Sacerdote (2001). De onderzoekers wilden kijken wat het winnen van een grote jaarlijkse geldprijs doet met het arbeidsmarktgedrag: gaan deze mensen minder werken? Zo ja, in welke mate? Het idee is dat de loterij een natuurlijk experiment is waarmee je het effect van een basisinkomen kan onderzoeken. Hierbij is de ‘interventie’ het winnen van een grote jaarlijkse geldprijs. Uit de analyse blijkt dat als het jaarlijkse prijzengeld rond de \$ 15.000 ligt (in dollars uit 2001), er vrijwel geen aanwijzingen zijn dat de prijswinnaars na het winnen van de prijs minder gaan werken dan ze voor het winnen van de prijs deden.

In de dataset is ook informatie beschikbaar over deelnemers die geen grote prijs hebben gewonnen. Diverse van deze deelnemers hebben weleens een prijs gewonnen in de loterij, maar dat is dan een relatief klein eenmalig bedrag. De informatie over de winnaars en niet-winnaars is afkomstig uit enquêtes. Hoewel je op voorhand zou verwachten dat winnaars van een loterij willekeurig vallen, blijkt dit niet het geval te zijn. Zo hebben deelnemers die meerdere loten kopen meer kans om te winnen. Maar er speelden ook andere factoren mee die maakten dat de groep winnaars en niet-winnaars niet zonder meer vergelijkbaar zijn. De groep winnaars die is bevestigd, bestaat uit vaste deelnemers en incidentele deelnemers. Voor de groep niet-winnaars is alleen informatie beschikbaar over vaste deelnemers, niet over incidentele deelnemers. Tenslotte is er sprake van selectieve non-respons op de enquête: naarmate mensen minder vaak (iets) hadden gewonnen in de loterij, was de geneigdheid kleiner om mee te doen aan de enquête. Daarom is de *propensity score* gebruikt om de twee groepen vergelijkbaar te maken. Er worden vervolgens verschillende gevoeligheidsanalyses uitgevoerd om na te gaan hoe robuust de uitkomsten zijn.

Dit voorbeeld laat zien dat zelfs in een relatief ‘eenvoudige’ setting van een loterij, waarbij geluk (en dus willekeur) een zeer grote rol speelt, er door verschillende oorzaken toch sprake kan zijn van twee groepen die niet zomaar met elkaar te vergelijken zijn.

Literatuur: Imbens (2015), Imbens, Rubin en Sacerdote (2001). Een handig overzichtsartikel vind je [hier](#).

Twee andere, nog niet eerder in rekenkameronderzoek gehanteerde, methoden in de context van een natuurlijk of quasi-experiment zijn een *regression discontinuity design* en de *difference-in-difference*-methode. Een *regression discontinuity design* kan gebruikt worden als er sprake is van een duidelijk afkappunt tussen wie er gebruik maakt van beleid en wie niet. Het idee van een *regression discontinuity* is dat de groep die net binnen de regeling of beleidsinterventie valt wordt vergeleken met de groep die net buiten de regeling of beleidsinterventie valt. Deze groepen lijken erg op elkaar maar het belangrijkste verschil is dat de ene groep wel gebruik heeft kunnen maken van de regeling en de andere niet. Vervolgens kan je voor deze groepen de grootte van het effect van de beleidsinterventie schatten. De beperking is dat deze methode alleen een uitspraak doet over het effect van het beleid rond het afkappunt.

Voorbeeld (hypothetisch): Onderwijsachterstanden

Om de onderwijsachterstanden te bestrijden, krijgen scholen waar meer dan 70% van de leerlingen een migratie-achtergrond heeft, een extra subsidie van de rijksoverheid. Belangrijke randvoorwaarde: deze regeling wordt strikt toegepast (geen uitzonderingen). Het bepalen van het percentage leerlingen met een migratie-achtergrond per school kan niet beïnvloed worden door de scholen. Vervolgens kan het effect van de extra subsidie voor het tegengaan van onderwijsachterstanden worden bepaald.

De *difference-in-difference*-methode is bruikbaar, als bijvoorbeeld het beleid in de tijd is veranderd. Hierdoor kan het effect bij een groep die onder de oude regeling viel vergeleken worden met het effect bij de groep die viel onder de nieuwe regeling.

Voorbeeld: Effect van beroepsonderwijs op inkomen

Een voorbeeld van toepassing van deze methode is het onderzoeken van het langetermijneffect op inkomen van de verlenging van de duur van het beroepsonderwijs in 1975. Voor 1975 duurde beroepsonderwijs drie of vier jaar. In 1975 werden alle driejarige opleidingen verlengd naar vier jaar en werd ook de leerplichtige leeftijd met een jaar verhoogd. Oosterbeek en Webbink (2004) hebben met behulp van de toepassing van de *difference-in-difference*-methode onderzocht welke langetermijneffecten deze verlenging heeft gehad op het inkomen van betrokkenen. Zij vonden dat de verlenging geen effect heeft gehad.

Literatuur: Hessel Oosterbeek en Dinand Webbink (2004). Wage effects of an extra year of lower vocational education: Evidence from a simultaneous change of compulsory school leaving age and program length. Working paper.

9.3.3 Analysemethoden voor onderzoek naar doelmatigheid

Naast analysemethoden voor het onderzoek naar doeltreffendheid bestaat er ook een aantal specifieke analysemethoden die kunnen worden ingezet om onderzoek te doen naar doelmatigheid. Hier wordt aan de hand van voorbeelden een aantal van dergelijke methoden besproken, namelijk benchmarking (9.3.3.1) en (ex post) maatschappelijke kosten-batenanalyse (9.3.3.2).

9.3.3.1 Benchmarking

Benchmarking is een analysemethode die ons helpt om te bepalen of beleid doelmatig is uitgevoerd. Door organisaties of instellingen die sterk op elkaar lijken (bijvoorbeeld ziekenhuizen, universiteiten of rechtbanken) met elkaar te vergelijken, kun je bepalen welke organisaties doelmatiger functioneren dan andere.

Doelmatigheid is hierbij dus een relatief begrip. Er zijn verschillende manieren om een benchmark uit te voeren: er kan worden gekeken naar specifieke processen, zoals de inkoopprocessen of wervings- en selectieprocessen. Maar organisaties kunnen ook vergeleken worden op prestaties, zoals het aantal behandelde patiënten of het aantal afgestudeerde studenten. Een relatief eenvoudige manier om eerste inzichten te verzamelen over doelmatigheid is door te kijken naar de kosten per product. Dit soort benchmarking van organisaties door te kijken naar kengetallen is een interessant startpunt bij het onderzoeken van doelmatigheid, maar er is meer nodig om te komen tot robuustere uitspraken over doelmatigheid.

De beperking van benchmarking op basis van kengetallen is dat er dan vaak geen rekening gehouden kan worden met verschillen in soorten productie (bijvoorbeeld eenvoudige behandeling versus ingewikkelde behandeling), verschillen in samenstelling van de populatie (bijvoorbeeld school in een achterstandswijk versus school in een villawijk) en met verschillen in kwaliteit van de diensten. Het verdient aanbeveling om hier zoveel mogelijk rekening mee te houden. Verder is het belangrijk om informatie per organisatie altijd in samenhang te bekijken: zo kan een organisatie heel goed scoren op een bepaald onderdeel (bijvoorbeeld er worden veel behandelingen verricht per medewerker), maar daar kan een slechte score tegenover staan op een ander onderdeel (bijvoorbeeld hoog ziekteverzuim). Het is dus belangrijk om bij de toepassing van benchmarking altijd oog te blijven houden voor het grotere geheel.

Voorbeeld

In het onderzoek [Hilversum in beeld](#) werden kosten van programma's van de publieke omroep met elkaar vergeleken. Hierbij werd onderscheid gemaakt tussen verschillende soorten programma's (zoals reisprogramma's, actualiteitenprogramma's en sport). Hieruit kwam naar voren dat bepaalde programma's relatief duur waren in vergelijking met gelijksoortige programma's. Dit kan een aanwijzing zijn voor ondoelmatigheid.

Een andere vorm van benchmarking is het vergelijken van regelingen die hetzelfde doel beogen, bijvoorbeeld het reduceren van de hoeveelheid CO₂-uitstoot. Door de kosten van een regeling te relateren aan het bereikte resultaat (in dit geval de kosten per bespaarde ton CO₂-uitstoot) is het mogelijk om te bepalen welke regeling het meest doelmatig is voor het behalen van beleidseffecten (kosteneffectiviteit). Deze regeling is dan doelmatiger dan de andere. Een beperking van deze aanpak is dat de regelingen alleen worden vergeleken op het effect dat wordt beoogd. Als de regelingen ook andere effecten hebben op de maatschappij (positief of negatief, bedoeld of onbedoeld), wordt dat in deze werkwijze niet meegenomen. Hierbij kan bijvoorbeeld gedacht worden aan geluidsoverlast van windturbines.

Voorbeeld

In het [Verantwoordingsonderzoek 2017](#) is onderzoek gedaan naar acht beleidsmaatregelen die worden ingezet om de luchtkwaliteit te verbeteren. Hierbij is gekeken naar het effect op twee onderdelen van luchtkwaliteit: emissiereductie (fijnstof) en stikstofdioxiden. Vervolgens is gekeken naar de kosteneffectiviteit van deze acht regelingen.

Een speciale vorm van benchmarking is een *productiviteitsanalyse*. Dit is een systematische wijze om te bepalen welke instellingen de hoogste verhouding tussen productie en ingezette middelen weten te bereiken. Daarbij wordt onder meer rekening gehouden met schaalgrootte. Op deze manier kun je vaststellen welke organisaties doelmatiger werken dan andere. Deze vorm van benchmarking kan wel expliciet rekening houden met zaken als verschillen in soort productie, samenstelling van de populatie en kwaliteit van de diensten, mits hier (voldoende) gegevens over beschikbaar zijn. Er zijn twee vormen van productiviteitsanalyse: *data envelopment analysis* (DEA) en *stochastic frontier analysis*. De DEA-methode is iets laagdrempeliger en hiervoor is standaardprogrammatuur beschikbaar. De beperking is dat er minder gedetailleerde uitspraken mogelijk zijn. Voor het toepassen van een *stochastic frontier analysis* moet de analyse op maat worden uitgevoerd. Dit vraagt de nodige kennis van econometrische modellen en van programmeren.

Een belangrijke beperking van de productiviteitsanalyse is dat je met toepassing van alleen deze methode niet boven water krijgt waarom sommige organisaties doelmatiger werken dan andere. Een andere praktische beperking is dat (systematische) gegevens over kwaliteit vaak niet aanwezig zijn, waardoor de methode geen onderscheid kan maken tussen organisaties die goede kwaliteit leveren en organisaties die dat niet doen. Dit kan ertoe leiden dat organisaties die een hoge productie leveren in combinatie met een slechte kwaliteit aangemerkt worden als doelmatig. Het is belangrijk hier rekening mee te houden in de uitvoering van het onderzoek en zoveel mogelijk aanvullende informatie te verzamelen over de geleverde kwaliteit.

Voorbeelden

In 2017 heeft de Zweedse rekenkamer onderzoek gedaan naar de doelmatigheid van [rechtbanken](#). In 2018 heeft de Deense rekenkamer onderzoek gedaan naar de doelmatigheid van [arbeidsbureaus](#)). Beide hebben zij hierbij een zogenaamde Data Envelopment Analysis (DEA) toegepast om de productiviteit te meten.

Literatuur: [Efficiency of Homes for the Mentally Disabled](#) van E. Eggink en J.L.T. Blank (Sociaal en Cultureel Planbureau, 2001) is een studie waarin de doelmatigheid van instellingen voor gehandicapten wordt onderzocht. In deze studie wordt zowel de *stochastic frontier analysis* als de DEA toegepast. De achtergrond van beide methoden wordt ook uitgebreid behandeld.

9.3.3.2 Maatschappelijke kosten-batenanalyse (ex post)

Een andere analysemethode om uitspraken te doen over de doelmatigheid of kosteneffectiviteit van beleid is de maatschappelijke kosten-batenanalyse (MKBA). In een MKBA staat doorgaans een bepaalde maatschappelijke opgave voor de overheid centraal. In een MKBA wordt gekeken welke beleidsmaatregel het meest doelmatig is om deze maatschappelijke opgave op te lossen. Hierbij wordt doelmatigheid onderzocht door de maatschappelijke baten die de beleidsmaatregel oplevert en de maatschappelijke kosten die daar tegenover staan met elkaar te vergelijken. Hierbij is het belangrijk op te merken dat ook deze methode ervan uitgaat dat doelmatigheid een relatief begrip is. In een MKBA worden verschillende beleidsopties onderzocht. In de regel is de variant met de beste verhouding tussen baten en kosten de meest doelmatige variant. Er is echter ook een ondergrens: als geen enkele variant een positieve verhouding laat zien tussen kosten en baten, is het niet doelmatig om dit beleid in uitvoering te nemen. Deze benadering van doelmatigheid is gestoeld op de welvaartseconomie en de bijbehorende sociale welvaartsfunctie. Met deze functie wordt de welvaart van

individuen opgeteld tot welvaart voor de hele maatschappij. Essentieel is dat in theorie de sociale welvaartsfunctie alles bevat wat voor mensen van belang is. Het idee van de MKBA is dat beleid niet alleen beoordeeld moet worden op de doelen die worden nagestreefd, maar expliciet op alle kosten en baten die dit beleid genereert. Als een beleids optie een positief saldo heeft, dan betekent dit dat het de welvaart verhoogt om deze beleids optie uit te voeren. Uiteraard kunnen er meerdere beleids opties zijn die welvaart verhogen. Daarom is het belangrijk om alle relevante beleidsvarianten mee te nemen in een MKBA om op die manier de meest doelmatige te identificeren.

Het is in Nederland verplicht om een MKBA uit te voeren voordat er een beslissing wordt genomen over een (grote) gebiedsveranderende investering waar de rijksoverheid aan meebetaalt. Deze MKBA vooraf noemen we ook wel een ex ante-MKBA.

De meeste MKBA's worden voorafgaand aan de implementatie van beleid uitgevoerd. Maar het ligt voor de hand om als rekenkamer juist achteraf (of wel ex post) te onderzoeken of de maatschappelijke baten van beleid opwegen tegen de maatschappelijke kosten. Voor het uitvoeren van een ex post-MKBA is het belangrijk dat er een ex ante-MKBA beschikbaar is. Als dit niet het geval is, zal het aanzienlijk meer tijd kosten om de MKBA uit te voeren. Ook zal het moeilijker zijn om bepaalde berekeningen uit te voeren, zeker als er geen informatie bekend is over bepaalde relevante effecten.

Bij een MKBA is het van belang om alle (belangrijke) kosten en baten te kunnen meten en om ze te kunnen uitdrukken in geld (monetariseren). Maar ook als dit niet mogelijk is, biedt het gedachtegoed van de MKBA een heel nuttig denkkader om beleidsmaatregelen te onderzoeken. In de algemene MKBA-leidraad, ontwikkeld door het CPB en het PBL, is een stappenplan voor een MKBA beschreven. Dit stappenplan bevat fundamentele vragen, zoals: hoe zou het probleem zich ontwikkelen zonder (overheids)ingrijpen? Welke alternatieven zijn er? Welke effecten (voorzien en onvoorzien) heeft elk van deze alternatieven op de maatschappij? Dit kader kan helpen de beleidstheorie te toetsen maar ook om onbedoelde effecten van beleid in kaart te brengen.

Voorbeeld

De Algemene Rekenkamer voerde een pilotproject uit waarin een ex post-MKBA wordt gedaan van een relatief klein project: de vervanging van een brug in het Burgemeester Delenkanaal. Hierbij werd in kaart gebracht of de maatschappelijke kosten en baten zoals die ex ante zijn becijferd ex post een goede weergave van maatschappelijke kosten en baten zijn geweest.

Literatuur: Algemene leidraad voor maatschappelijke kosten-batenanalyse, Gerbert Romijn en Gusta Renes, CPB/PBL 2013.

9.3.4 Verder lezen over analysemethoden

In dit hoofdstuk is slechts een beperkt aantal analysemethoden aan de orde gekomen. Het palet in te zetten analysemethoden voor rekenkameronderzoek is vele malen groter. De volgende websites kunnen ter inspiratie worden geraadpleegd bij het zoeken naar analysemethoden die passen bij de onderzoeksvragen in een specifiek rekenkameronderzoek:

- De [Toolbox](#)-beleidsevaluatie, ontwikkeld in opdracht van het ministerie van Financiën, is een nuttig hulpmiddel bij het maken van het onderzoeksdesign. Daar vind je informatie over verschillende onderzoeksmethoden die aansluiten bij verschillende soorten onderzoeksvragen.
- Het [Rainbow Framework](#), ontwikkeld door de non-profitorganisatie Better Evaluation, bevat een uitgebreid overzicht van mogelijke methoden. Zij benaderen de methoden ook vanuit verschillende soorten onderzoeksvragen: beschrijven (*describe*), causale verbanden onderzoeken (*understand causes*), overkoepelende conclusies trekken (*synthesize*) en aanbevelingen formuleren (*report & support use*).
- Het [Handbook Impact Evaluation in Practice](#) van de Wereldbank geeft een overzicht van diverse methoden voor effectevaluatie.

Literatuur

Austin P. C. (2011). An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate behavioral research*, 46(3), 399–424.

De Graaf, Djoerd, Lucy Kok, Peter Berkhout (2004). *Is het mogelijk integratie te meten?* SEO: Amsterdam.

Eggink, E., J.L.T. Blank (2001). *Efficiency of Homes for the Mentally Disabled*. Den Haag: Sociaal en Cultureel Planbureau.

Hauser, John R., Vithala R. Rao (2002). *Conjoint Analysis, Related Modeling, and Application*. In: Wind Y., Green P.E. (eds) *Marketing Research and Modeling: Progress and Prospects*. International Series in Quantitative Marketing, vol 14. Springer, Boston, MA.

Imbens, Guido W., Donald B. Rubin, Bruce I. Sacerdote (2001). Estimating the effect of unearned income on labor earnings, savings, and consumption: Evidence from a survey of lottery players. *American Economic Review*, Vol. 91, Issue 4, pages 778-794.

Imbens, Guido W. (2015). Matching methods in practice: Three examples. *Journal of Human Resources*, Volume 50, Issue 2, pages 373-419

Klein Haarhuis, C.M., E. Niemijer (2008). *Wet en werkelijkheid. Bevindingen uit evaluaties van wetten. Onderzoek en beleid 267*. WODC: Den Haag.

Robson, Colin (2002). *Real world research*. Blackwell Publishing.

Romijn, Gerbert, Gusta Renes (2013). *Algemene leidraad voor maatschappelijke kosten-batenanalyse*. Den Haag: CPB/PBL.

Swanborn, P.G. (1999). *Evalueren*. Meppel: Boom.

Van der Knaap, L., S. Bogaerts (2006). *Research synthese Geweld verslagen. Factsheets 2006-23*. Den Haag: WODC.

Van Praag, B.M.S. en P. Frijters (1997). *Choice behaviour and verbal behaviour: a critical assessment of their relevance for practical policy. A presentation for the 10th anniversary Tinbergen Conference, Empirical models and policy making. Also Tinbergen Discussion Paper 97-119/1*.

Veenma, K., R. Batenburg & E. Breedveld (2004). *De Vignetmethode. Een praktische handreiking bij beleidsonderzoek*. Tilburg: IVA.

Verweij & Gerrits (2012). *Systematisch kwalitatief vergelijkende analyse*. KWALON – Tijdschrift voor Kwalitatief Onderzoek, 17(3), 25-33.

Wartna, B.S.J., D.L. Alberda, S. Verweij (2013). *Wat werkt in Nederland en wat niet? Onderzoek en beleid 307*. Den Haag: WODC.

10.

Bevindingen, conclusies en bestuurlijke uitspraken

10.1 Inleiding

Het formuleren van bevindingen, conclusies, oordelen en aanbevelingen is een belangrijk onderdeel van het onderzoek. Ze bepalen in hoge mate de impact ervan. Als je voldoende en gepast bewijsmateriaal hebt verzameld, heb je een goede basis om tot bevindingen te komen en conclusies te trekken, zodat je de doelstelling van het onderzoek bereikt, de onderzoeksvragen kunt beantwoorden en voorstellen voor oordelen en aanbevelingen kunt formuleren.

Het is belangrijk om de begrippen bevindingen, conclusies, oordelen en aanbevelingen goed te onderscheiden en er inhoud aan te geven. We zullen daarom afzonderlijk op deze begrippen ingaan en deze ook illustreren en toelichten met voorbeelden, valkuilen en tips.

10.2 Bevindingen

In veel rekenkameronderzoeken stellen we toetsende (normatieve) vragen. Het kenmerkende van een toetsende vraag is dat je een of meer normen (de gewenste situatie) en de praktijk (de werkelijke situatie) met elkaar vergelijkt (zie ook hoofdstuk 05). Je bepaalt zoveel mogelijk vooraf welke normen gelden. Deze neem je op in het projectvoorstel. Voorbeeld: een onderzoek naar wachttijden in de jeugdzorg. De norm kan dan zijn: de maximale wachttijden die in wet- en regelgeving zijn vastgelegd.

Tijdens het onderzoek verzamel je onderbouwde feiten – bewijsmateriaal – over de werkelijke situatie en die vergelijk je met de norm. Het resultaat van die vergelijking noemen we een bevinding. Voorbeeld: wij constateren dat in de periode 2015-2018

voor het onderdeel opvoedhulp in 35% van de gevallen de wachttijd is overschreden. Dat is een constatering, dus zonder waardeoordeel.

Op zichzelf is het mogelijk je te beperken tot dergelijke constateringen en zo je onderzoeksvragen te beantwoorden, maar je kunt de toegevoegde waarde van je onderzoek aanzienlijk vergroten door het stellen van twee aanvullende vragen, namelijk naar oorzaken en gevolgen van je bevinding.

Let op: in veel gevallen heb je hier bij het opstellen van je projectvoorstel nog geen zicht op. Met name het zoeken naar verklaringen kan ingewikkeld en tijdrovend zijn. Als tijdens je onderzoek blijkt dat je hier niet genoeg tijd voor hebt uitgetrokken, moet je het projectvoorstel (contract met het college) aanpassen en daarvoor toestemming vragen van directeur en college.

- Wat is de (mogelijke) oorzaak van onze bevinding?

Met deze vraag willen we nagaan waarom de werkelijkheid afwijkt van de norm. Voorbeeld: met de constatering dat de wachttijd voor opvoedhulp in 35% van de gevallen is overschreden maak je duidelijk dat er een probleem is, maar als je niet weet wat de oorzaak is, weet je ook niet hoe dit probleem kan worden opgelost. Nader onderzoek zou kunnen uitwijzen dat de hulp niet goed georganiseerd is of dat de hulptrajecten veel langer duren dan voorzien was. Je kunt ook nog verder zoeken naar oorzaken van de lange duur van hulptrajecten en tot de conclusie komen dat hulpverleners in ernstige gevallen niet tijdig doorverwijzen naar bijvoorbeeld de jeugdbescherming. Zo kom je steeds verder tot de kern van het probleem (de *root cause*) en kun je ook veel gerichtere aanbevelingen doen.

Bij het onderzoek naar oorzaken is het belangrijk dat je je breed oriënteert.

Raadpleeg daarom ook hoofdstuk 06. Oorzaken kunnen bijvoorbeeld gerelateerd zijn aan de organisatie (bedrijfsvoering en governance), aan de beschikbare middelen (geld en capaciteit), maar ook aan minder 'harde' aspecten, zoals cultuur en gedrag (pilot).

De vraag naar oorzaken behoort tot de categorie verklarende onderzoeksvragen.

- Wat is het (mogelijke) gevolg van onze bevinding?

Met deze vraag willen we nagaan hoe erg het is dat de werkelijkheid afwijkt van de norm. De bedoeling is dat we hiermee ook een gevoel van urgentie creëren en daarmee input leveren voor een bestuurlijk oordeel.

Bij deze vraag is het belangrijk om, voor zover van toepassing, te kijken vanuit het perspectief van de burger: in hoeverre is het probleem schadelijk voor degenen die afhankelijk zijn van het beleid? Nader onderzoek naar de gevolgen van de overschrijding van de wachttijd voor opvoedhulp zou kunnen uitwijzen dat dit in veel

gevallen leidt tot ontsporingen in gezinnen, waardoor zwaardere (en duurdere) hulpverleningstrajecten moeten worden ingezet.

De vraag naar gevolgen kan zowel beschrijvend als toetsend zijn. Bij een beschrijvende vraag breng je bijvoorbeeld mogelijke gevolgen in kaart en bij een toetsende vraag stel je bijvoorbeeld als norm dat er geen sprake mag zijn van ongewenste (neven)effecten van het beleid.

10.3 Feitelijke conclusies

In ons onderzoek geven we uiteindelijk antwoord op de onderzoeksvragen. Het antwoord op een beschrijvende onderzoeksvraag is feitelijk beschrijvend. Er zijn geen bevindingen, omdat we geen normen hebben gehanteerd. Je moet dan niet spreken van bevindingen of conclusies, maar bijvoorbeeld van observaties. Het antwoord op een toetsende onderzoeksvraag is een feitelijke conclusie, dus zonder waardeoordeel. Het antwoord op een verklarende onderzoeksvraag kan zowel beschrijvend als concluderend zijn.

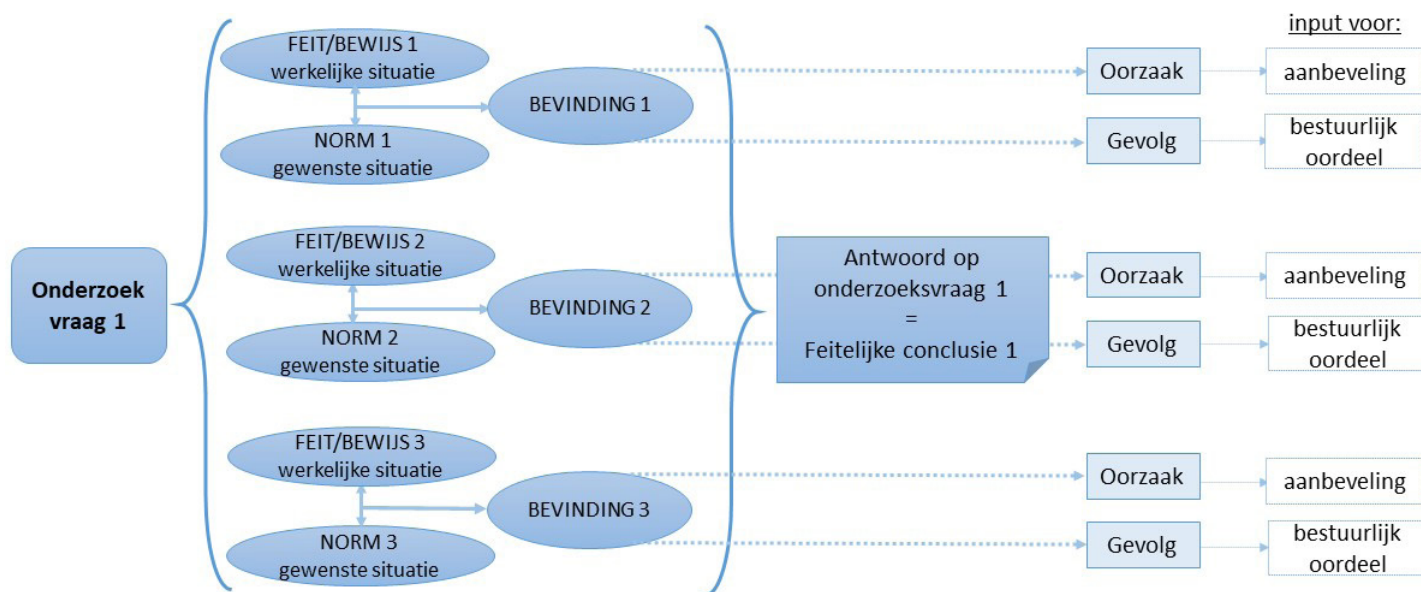
In het algemeen zijn meer bevindingen nodig om een onderzoeksvraag te beantwoorden. In het voorbeeld van de wachttijden ging het om een onderzoek naar 12 wachttijden voor verschillende vormen van jeugdzorg. Per wachttijd heb je dan een bevinding. De feitelijke conclusie is het antwoord dat geldt voor alle bevindingen samen. Ook dan is je antwoord feitelijk, dus zonder waardeoordeel. Een mogelijke feitelijke conclusie zou kunnen zijn: de minister voldoet niet aan alle maximale wachttijden in de jeugdzorg. We stellen vast dat 8 van de 12 wachttijden in de periode 2016-2019 zijn overschreden met gemiddeld 15%.

10.4 Bevindingen en feitelijke conclusies: de nota van bevindingen

De nota van bevindingen is een typisch rekenkamerproduct. Het is een weerslag van het bewijsmateriaal (de feiten) die we met ons onderzoek hebben verzameld in relatie tot de normen die we vooraf hebben gesteld aan de door ons onderzochte werkelijkheid.

De nota van bevindingen is vormvrij, maar moet wel een heldere structuur hebben, bij voorkeur geordend volgens de onderzoeksvragen. Als de nota van bevindingen niet volgens de onderzoeksvragen is gestructureerd, moet je afsluiten met een hoofdstuk waarin deze vragen expliciet worden beantwoord. In schema 10.1 vind je een handvat voor het ordenen van de nota van bevindingen en alle elementen die daarbij een rol spelen.

Schema 10.1 Een structuur voor de nota van bevindingen



In het schema is één onderzoeksvraag uitgewerkt, maar voor de andere onderzoeksvragen geldt uiteraard dezelfde structuur.

De nota moet duidelijk zijn en geen overbodige informatie bevatten, om te voorkomen dat de gecontroleerde deze overbodige informatie ook moet beoordelen tijdens de fase van ambtelijk wederhoor. Dat betekent dat je in de nota van bevindingen alleen relevante informatie opneemt. Behalve het volgen van een heldere structuur die is afgeleid van de onderzoeksvragen, kunnen ook de volgende tips hierbij helpen:

- Zorg ervoor dat de onderzoeksbevindingen die je in de nota opneemt, expliciet bijdragen aan de beantwoording van de onderzoeksvragen en de probleemstelling. Dat vergt een scherp onderscheid tussen ‘nice to know’ en ‘need to know’.
- Vermeld bij de beschrijving van de werkelijke situatie de gehanteerde bronnen. Dan is het niet nodig om uitgebreid te citeren uit brondocumenten.

Voorbeeld: Je onderzoeksvraag luidt: “Heeft de minister de aard en omvang van het tekort vanaf 2018 voldoende in beeld gehad?”

- Als start van het onderzoek naar deze vraag bepaal je wat de minister in beeld had moeten hebben. Dit zijn de normen waaraan je gaat toetsen.
- Vervolgens verzamel je de benodigde informatie.
- In de nota van bevindingen beschrijf je of aan de normen voldaan is: “de minister heeft dit in beeld, dit deels en dit niet”. Je kunt de beantwoording van deze onderzoeksvraag in de nota van bevindingen afronden (of beginnen) met een feitelijke conclusie: “de minister heeft de aard en omvang van het tekort gedeeltelijk in beeld gehad”.

10.5 Bestuurlijke uitspraken

In een eindrapport of publicatie beperken rekenkamers zich meestal niet tot feitelijke conclusies. In de publicaties doen we namelijk vaak bestuurlijke uitspraken, die we ook wel bestuurlijke boodschap noemen. Bestuurlijke uitspraken kunnen we onderscheiden naar oordelen, aanbevelingen en overige uitspraken.

Een bestuurlijk oordeel is een waardering van de bevindingen en de feitelijke conclusies van ons onderzoek. Dit is een waardeoordeel, dat aangeeft hoe erg de rekenkamer het vindt dat we – via een bevinding – vaststellen dat een situatie in de praktijk afwijkt van de norm of van de verwachte situatie. Zo'n oordeel is een kwalificatie als: onvoldoende, tekortschietend, onjuist, te laat, te duur of te weinig. Voor deze kwalificaties hebben we geen normen, behalve dat bestuurlijke oordelen niet strijdig mogen zijn met de feitelijke conclusies. In het voorbeeld zou een bestuurlijk oordeel kunnen zijn: De wachttijden in de jeugdzorg worden door de minister onvoldoende beheerst.

Een bestuurlijke aanbeveling vloeit voort uit een bestuurlijk oordeel. Aanbevelingen zijn belangrijke instrumenten om de missie van de rekenkamer (bijdragen aan het verbeteren van het functioneren en presteren van de rijksoverheid en de daarmee verbonden organen) te realiseren.

Een goede aanbeveling:

- wordt gedragen door de bevindingen en conclusie(s) van het onderzoek. Een aanbeveling die niet bij een bevinding of conclusie uit het onderzoek past, kun je opnemen, maar dan begin je met "Overigens zijn wij van mening dat..." Of "Aanvullend vinden wij dat...";
- is consistent met de andere aanbevelingen in het product en in principe consistent met eerdere aanbevelingen van de rekenkamer op het desbetreffende terrein. Maar vanzelfsprekend is een beargumenteerde afwijking mogelijk. Hiervoor moet het team dan wel een onderbouwing hebben, die in het dossier aanwezig moet zijn (hoeft niet in het product zelf);
- is probleemoplossend. De aanbeveling moet kunnen bijdragen aan het oplossen van de gesignaleerde problemen. Het is daarom belangrijk dat aanbevelingen aangrijpen bij oorzaken van tekortkomingen en problemen. Zie hiervoor ook 10.02. De aanbeveling is zo geformuleerd dat de aangesprokene moet werken aan de oplossing van het gesignaleerde probleem. Bij voorkeur dus geen procesaanbeveling zoals "bezie", "overweeg" of "overleg" als hoofdbestanddeel van de aanbeveling. Als de aanbeveling zich toch op het proces richt, dan moet altijd duidelijk zijn waarop die procesaanbeveling gericht is ("doe x, zodat...");
- heeft toegevoegde waarde. Aanbevelingen doen is niet verplicht; bij twijfel geen aanbeveling;

- is actorgericht. Het moet duidelijk zijn wie iets moet doen en die 'wie' moet de bevoegdheid hebben iets aan de oplossing van het probleem te doen;
- is uitvoerbaar. De aanbeveling mag alleen een 'hoe' bevatten als er onderbouwing is – vanuit het onderzoek zelf of uit andere bronnen – dat deze oplossing werkt;
- is helder geformuleerd. De aanbeveling moet begrijpelijk zijn geformuleerd (geen vaag taalgebruik) en voldoende concreet zijn. Het moet duidelijk zijn dat het om een aanbeveling gaat (dus geen 'aandachtspunt', 'overweging' of iets dergelijks). De (opvolging van) de aanbeveling is meetbaar en toetsbaar. Een aanbeveling mag een omgekeerde conclusie zijn, als er geen beter alternatief is.

Vraag je bij het formuleren van aanbevelingen ook altijd af of de voorgestelde verbeteringen geen ongewenste (neven)effecten kunnen hebben.

Bij *overige bestuurlijke uitspraken* kun je bijvoorbeeld denken aan het uitspreken van verwachtingen ("we gaan ervan uit dat..."), van voornemens ("we zullen de opvolging van de toezeggingen met aandacht blijven volgen...") of van begrip.

Tabel 10.2 *Samenhang tussen nota van bevindingen en eindrapport*

Nota van bevindingen	Eindrapport	Bestuurlijke boodschap
Normen	Relevante normen	<i>Relevant voor de onderbouwing van de bestuurlijke boodschap</i>
Feiten/bewijs	Relevante feiten/bewijs	
Bevindingen	Relevante bevindingen	
Conclusies (antwoorden op onderzoeksvragen)	Relevante conclusies	
	Oordelen Aanbevelingen Overige bestuurlijke uitspraken	<i>Bestuurlijke boodschap</i>

11.

Methodische bijlage

11.1 Inleiding

Voor het vertrouwen in rekenkameronderzoek is het belangrijk om uit te leggen hoe we tot onze bevindingen en conclusies zijn gekomen. Een belangrijk aspect daarbij is dat navolgbaar moet zijn hoe wij ons onderzoek hebben uitgevoerd en dat wij in staat zijn onze methodische en methodologische keuzes te verantwoorden. Daarbij is een methodische verantwoording bij het onderzoeksrapport in een methodische bijlage een belangrijk instrument. In de methodische bijlage geven we aan hoe we het onderzoek hebben uitgevoerd, welke keuzes we gemaakt hebben en wat de beperkingen van deze keuzes zijn. We leggen uit hoe we tot onze bevindingen en conclusies zijn gekomen. Het doel van de methodische bijlage is om transparant uitleg te geven over het onderzoek, zodat lezers zich een goed beeld kunnen vormen van de reikwijdte, beperkingen en betrouwbaarheid van het onderzoek. Het werken met een methodische bijlage past ook bij naleving van internationale controlestandaarden voor rekenkamers, de ISSAIs (<https://www.issai.org>), die onderdeel uitmaken van het IFPP. De methodische bijlage zorgt ervoor dat duidelijk is voor de lezer hoe we tot onze uitkomsten zijn gekomen zonder de lezer te vermoeien met details hierover in de hoofdtekst.

De methodische bijlage is geen recept; een lezer hoeft niet op basis van de methodische bijlage het onderzoek over te kunnen doen. Replicatie is vaak ook niet mogelijk, omdat anderen niet dezelfde bevoegdheid hebben, dus niet bij dezelfde informatie kunnen komen. Aan de andere kant moet een methodische bijlage verder gaan dan “we hebben dossieronderzoek gedaan” of “we hebben data geanalyseerd”. Tussen die twee uitersten moeten onderzoekers keuzes maken over wat ze wel en

niet opnemen in de methodische bijlage en hoe gedetailleerd ze dit doen. Hierbij geldt dat het doel van transparantie leidend is.

Het schrijven van een goede methodische bijlage is dus maatwerk. De omvang van de bijlage zal afhangen van de eigenschappen van het onderzoek. In uitgebreidere en complexe onderzoeken, waarbij er diverse onderzoeksmethoden zijn gebruikt (bijvoorbeeld ingewikkelde data-analyses met vele databestanden in combinatie met gestructureerd dossieronderzoek, een geavanceerde enquête en een observatieonderzoek) is het advies een uitgebreide methodische bijlage te schrijven.

Wat moet er in de methodische bijlage staan?

Bij de invulling van de methodische bijlage is dus vrij veel ruimte, maar deze moet in elk geval voldoende informatie geven over:

- de doelstelling, probleemstelling en onderzoeksvragen;
- de gehanteerde normen (voor toetsende onderzoeken);
- de methode en de daarbij gebruikte data en de wijze waarop die data zijn gebruikt om antwoord te geven op de onderzoeksvragen. Het verdient sterke aanbeveling om dit per onderzoeksvraag te beschrijven.

Een goede voorbereiding is het halve werk

Je kunt alleen een goede methodologische verantwoording schrijven, als je van tevoren goed hebt nagedacht over de keuzes die je maakt in je onderzoek. Het schrijven van een goede methodologische verantwoording begint dus al bij de keuze van de opzet, data, en analysemethoden van je onderzoek tijdens het schrijven van de strategische startnotitie of het projectvoorstel. Schrijf de methodische bijlage dus niet helemaal aan het eind van je onderzoek.

Meer dan tekst alleen

Voor alle onderdelen in de methodische bijlage geldt: beeld is fijn! Een ingewikkelde analyse uitleggen kan vaak beter met een figuur waarin we de samenhang der dingen uitleggen.

11.2 Tips en voorbeelden bij het schrijven van een methodische bijlage

Hier vind je tips hoe je invulling kunt geven aan de verschillende onderdelen van een methodische bijlage. Hierbij zijn ter inspiratie diverse voorbeelden opgenomen.

Let op: de methodische bijlage moet los van het rapport te begrijpen zijn.

11.2.1 Wat hebben we onderzocht?

- beschrijving van het object van onderzoek: bijvoorbeeld onderwerp, beleids-terrein, regeling, onderzochte entiteiten en periode waarop het onderzoek betrekking heeft;
- doelstelling en probleemstelling;
- onderzoeksvragen.

Voorbeeld: onderzoek [Paardenmiddel of noodverband? Resultaten prijsonderhandelingen geneesmiddelen](#)

Met dit onderzoek wil de Algemene Rekenkamer inzicht bieden in de vraag in hoeverre de door het Ministerie van VWS gevoerde prijsonderhandelingen hebben bijgedragen aan het bereiken van kosteneffectieve zorg en de beheersing van de geneesmiddelenuitgaven. Daarbij hanteren wij het uitgangspunt dat bij de adviesprijs van het zorginstituut sprake is van kosteneffectieve zorg. We hebben onderzocht:

- of het Ministerie van VWS er in de prijsonderhandelingen met fabrikanten in slaagt de adviesprijs van het Zorginstituut te bedingen;
- wat de uitgaven aan de geneesmiddelen zouden zijn geweest wanneer voor elk geneesmiddel de adviesprijs zou zijn betaald.

We hebben dit onderzocht voor alle 32 arrangementen gesloten in de periode 2012-2018.

11.2.2 Gebruikte normen

In een audit wordt getoetst aan normen, wat ook betekent dat er een toetsende onderzoeksvraag wordt gebruikt. Het bekendste voorbeeld van een 'non-audit' zijn focusonderzoeken, waarbij alleen feiten op een rij worden gezet zonder een norm te toetsen en zonder een bestuurlijk oordeel te geven. Bij elk onderzoek dient de methodische bijlage daarom een passage te bevatten die het voor de lezer duidelijk maakt wat voor soort onderzoek het betreft. Als er sprake is van toetsend onderzoek, moet je de gehanteerde normen opnemen in de methodische bijlage.

11.2.3 Aanpak van het onderzoek

11.2.3.1 Beschrijving van de methode

In elk onderzoek worden een of meer methoden gehanteerd waarmee de onderzoeksvragen worden beantwoord. Geef hier een hoog-over-omschrijving van wat we gedaan hebben. Gebruik bijvoorbeeld de wetenschappelijke of officiële benaming van de methoden, en verwijfs zo nodig naar een bron die de methoden beschrijft.

Voorbeeld: onderzoek Handhaven in het duister

Onderscheid tussen BRZO en net-niet-BRZO

Het onderzoek richt zich op bedrijven die vallen onder het Besluit risico's zware ongevallen (BRZO) en bedrijven die hier net buiten vallen ('net-niet-BRZO'). De 482 bedrijfslocaties die we hebben onderzocht, hebben we in twee groepen ingedeeld: BRZO-bedrijfslocaties en net-niet-BRZO-bedrijfslocaties. Bedrijfslocaties die in de periode 2015-2019 de meerderheid van de jaren (peildatum 1 maart van het desbetreffende jaar) de BRZO-status hadden, kregen van ons het kenmerk BRZO, de rest was net-niet-BRZO. Zo krijgen 351 bedrijfslocaties van ons het label 'BRZO-bedrijf' en 131 bedrijfslocaties merken we aan als 'net-niet-BRZO-bedrijven'.

Vergelijking regio's

Bij de vergelijking van de regio's is het van belang aan te geven dat we hier de postcode van de bedrijfslocaties als uitgangspunt nemen. Op basis van de data van Regioatlas hebben we elke postcode gekoppeld aan 1 van de 29 'gewone' omgevingsdiensten en aan 1 van de 6 BRZO-omgevingsdiensten.

Analyse van zwaarte van de handhaving

Voor de analyse op het eind van § 5.3 (wordt de handhaving uitgevoerd volgens de landelijke handhavingsstrategie?) hebben we een *generalized linear model* gebruikt. Hierin houden we rekening met de geneste structuur van de data: inspecties-met-overtredingen binnen bedrijven. De zwaarte van de handhaving is hierbij ingedeeld in vier categorieën, van licht naar zwaar: (1) hercontrole, informeren, (2) waarschuwen, LOD, LOB, (3) stilleggen, bestuurlijke boete, bestuurlijke strafbeschikking, (4) proces-verbaal, strafrecht. Omdat de te verklaren variabele een ordinaal meetniveau heeft (een vierpuntsindeling van handhavingszwaarte) hebben we zogeheten ordinale logistische regressie toegepast.

De methode bepaalt en beperkt eigenlijk altijd het soort uitspraken dat je kunt doen (zie ook 01.03). Het kan zijn dat je onderzoek een analyse is van alle data over een bepaalde regeling binnen een bepaalde periode (bijvoorbeeld ouderdomsregelingen, landbouwsubsidies, toeslagen). Het kan ook zijn dat er wordt gebruikgemaakt van een representatieve steekproef. Soms worden bepaalde casussen gekozen die een aantal variaties van verschijnselen kunnen laten zien. Soms kunnen casussen zo worden gekozen dat ze het alsnog mogelijk maken om bredere uitspraken te doen. Het is daarom belangrijk dat in de methodische bijlage wordt beschreven wat de methode van het onderzoek is, en welke beperkingen deze methode kent. Welke zeggingskracht heeft het onderzoek over de periode buiten de geanalyseerde periode? Welke groepen zijn minder vertegenwoordigd in de steekproef en in

hoeverre zijn daar oplossingen voor? In hoeverre heeft dit onderzoek zeggingskracht over niet onderzochte casussen? Geef ook aandacht aan de implicaties hiervan voor de conclusies.

Voorbeeld: onderzoek Inzicht in kwaliteit, operatie geslaagd?

In dit onderzoek hebben we een selectie van negen initiatieven gemaakt. We gingen daarbij uit van de voorwaarde dat we initiatieven van negen verschillende departementen wilden bekijken. Vervolgens hebben we bij de keuze geprobeerd zo veel mogelijk variatie te bereiken in indeling, duur, type en focus. Op die manier genereren we een zo goed mogelijk beeld van de verschillende soorten initiatieven. Het is dan ook te verwachten dat onze inzichten grotendeels ook zullen opgaan voor de niet onderzochte initiatieven.

Voor elk onderdeel van het onderzoek leggen we de aanpak uit. Als er voor één of meer onderwerpen meer dan één methode is gebruikt, lichten we alle methoden toe.

11.2.3.2 Beschrijving van de data

De technische definitie van data in onderzoek is “alle informatie die verzameld, geobserveerd, gegenereerd of gemaakt is om bevindingen te ondersteunen en valideren. Dit kunnen dus Excel-bestanden met cijfers zijn, maar ook gespreksverslagen, observaties of documenten. We doen soms ook literatuuronderzoek, dit beschouwen we meestal niet als data, omdat we dit meer als achtergrond voor bevindingen gebruiken. Als we een systematische literatuurreview doen, kunnen we literatuur ook als data beschouwen.

Let op: data zijn dus alle informatie die we gebruiken om bevindingen op te baseren.

Gebruikte data

In de methodische bijlage beschrijven we per onderzoeksvraag eerst de data die we geanalyseerd hebben:

- Welk type data? (Bijvoorbeeld: Excel-bestand, gespreksverslagen, beleidsdocumenten of een combinatie daarvan).
- Wat is de bron van de data?
- Hebben we de data zelf verzameld, of zijn ze aangeleverd door de gecontroleerde?
- Als er voor de bevindingen is gebruikgemaakt van (uitkomsten uit) onderzoeksrapporten die door anderen zijn uitgevoerd, hoe is de kwaliteit van deze rapporten dan gevalideerd (zie hiervoor ook 08.06)?
- Als de data zijn aangeleverd door de gecontroleerde: wat weten we over de kwaliteit van de data en wat betekent dat voor ons onderzoek? Hoe hebben we

bijvoorbeeld de kwaliteit van de data gecontroleerd en vastgesteld (zie hiervoor ook 08.05)?

- Als we de data zelf hebben verzameld: op welke manier hebben we dit gedaan, en wat waren de mogelijke beperkingen van deze methode?
- Hebben de data bepaalde beperkingen die van belang zijn voor het beantwoorden van de onderzoeksvragen en daarmee voor de reikwijdte en/of generaliseerbaarheid van de conclusies?
- Als er sprake is van gegevens over personen of bedrijven: hoe hebben we de privacy gewaarborgd?

Voorbeeld: onderzoek Sociale zekerheid en flexibele arbeidsmarkt

We hebben microdata van het Centraal Bureau voor de Statistiek (CBS) gebruikt. Onze data-analyse is gebaseerd op de volgende databestanden:

- Sociaaleconomische categorieën van personen (SECMBUS);
- Persoonskenmerken van alle personen in de Gemeentelijke basisadministratie persoonsgegevens (GBAPERSONTAB);
- Huishoudenskenmerken (GBAHUISHOUDENSBUS);
- Opleidingsgegevens van personen (HOOGSTEOPLTAB);
- Banen en lonen op basis van de polisadministratie (SPOLISBUS 2010 t/m 2017);
- Inkomen van zelfstandigen (ZELFSTANDIGENMNCBDRAGBUS)
- Inkomensgegevens, waaronder pensioen (INPATAB);
- Vermogens van huishoudens (VEHTAB).

Wij hebben data van de jaren 2010 tot en met 2017 gebruikt. Voor persoons- en huishoudensgegevens zoals leeftijd en plaats in het huishouden hebben we het peilmoment 1 januari 2017 gekozen.

[...]

Beperkingen data

- De geraadpleegde CBS-microdata zijn inkomens op jaarbasis. Binnen een jaar kan het inkomen echter verschillen of zelfs wegvallen. Met onze gemiddelden, alles opgeteld en gedeeld door de meetperiode van een persoon, is er de mogelijkheid op onderschatting of overschatting van het inkomen.
- Bij het koppelen van gegevens van de partner, zijn alleen twee partners die samen een huishouden vormen meegenomen. Bij huishoudens met meer dan twee personen die als partner aangemerkt zijn, was deze koppeling niet mogelijk omdat we niet konden vaststellen wie partners van elkaar zijn.

Voorbeeld: onderzoek Handhaven in het duister

De analyses in dit rapport konden pas gedaan worden na vele kwaliteitsslagen op de data. De problemen die we aantreffen in de kwaliteit van de data waren zo ernstig dat we hier een apart rapport over gepubliceerd hebben (Een onzichtbaar probleem. Gebrek aan kwaliteit van data over milieucriminaliteit en -overtredingen.)

We hebben voor dit onderzoek een aantal databronnen gebruikt (op volgorde van relevantie voor dit onderzoek):

- Inspectieview;
- data van zes BRZO-omgevingsdiensten;
- justitiële data uit JDS;
- data van de Kamer van Koophandel, via dataportal van Company.info;
- bedrijvenlijsten Gemeenschappelijke Inspectieruimte (GIR), aangeleverd door Rijkswaterstaat (RWS);
- GIR: door universiteiten Leiden en Amsterdam omgezet in datavorm.

Daarnaast hebben we nog hulpdatabestanden van bijvoorbeeld het CBS gebruikt (voor postcode-gemeente-indelingen en voor *shapefiles* voor een geografische figuur).

De grootste uitdaging in dit project was om tussen en binnen al deze databestanden de juiste data aan het juiste bedrijf te koppelen. Daarnaast hadden we te kampen met veel ontbrekende gegevens. In de volgende alinea's gaan we hierop in.

Voorbeeld: onderzoek Toeslagen terugbetalen

Omdat het hier persoonsgegevens betreft, heeft een privacy impact assessment plaatsgevonden. De gegevensverwerking van dit onderzoek is geregistreerd in het AVG-register van de Algemene Rekenkamer.

Databewerking en -verwerking

- Welke bewerkingen hebben we op de ontvangen 'ruwe' data gedaan voordat we deze hebben geanalyseerd? Bijvoorbeeld: hoe hebben we verschillende datasets gecombineerd of hoe hebben we de data geaggregeerd?
- Welke variabelen/constructen/features hebben we op basis van de ontvangen data zelf aangemaakt en wat is de definitie hiervan?
- Welke bewerkingsslagen hebben we toegepast op de data en welke mogelijke gevolgen heeft dit voor de conclusies? Bij kwantitatieve data gaat het bijvoorbeeld om koppelingen tussen bestanden en omgaan met extreme

waarden of ontbrekende waarden. Bij kwalitatieve data gaat het bijvoorbeeld om de vraag of er deductief of inductief is gecodeerd (dus met een vooraf bepaald codeschema of niet). Is er gekeken naar veelvoorkomende thema's, is er geturfd?

Voorbeeld: onderzoek Handhaven in het duister

In het onderzoek Handhaven in het duister zijn diverse databronnen gebruikt, waarbij de grootste uitdaging lag in het koppelen van het juiste bedrijf aan de juiste gegevens. In de methodische bijlage wordt hierbij uitgebreid stilgestaan. Hieronder volgen een paar passages:

“Omdat de spelling van bedrijfsnamen en de adressen/adresformuleringen in de BRZO-lijsten over de verschillende jaren niet stabiel waren, hebben we handmatig gecontroleerd of een bedrijfslocatie in het ene jaar identiek was aan een bedrijfslocatie in het andere jaar. Hiervoor hebben we soms ook gebruik gemaakt van extra openbare bronnen, zoals de portal van Company.info.”

“We hebben de koppelingen op diverse manieren gelegd. Steeds in eerste instantie via automatische koppelingen (op basis van bedrijfsnamen, adresgegevens en Kamer van Koophandel-nummers). Waar nodig hebben we ook een meer handmatige aanpak gekozen. Zo hebben we binnen postcodegebieden handmatig/visueel gezocht naar bedrijfsnamen die op elkaar leken (of waarvan we wisten dat er naamswijzigingen plaats hadden gevonden, denk aan Hoogovens/Tata Steel en Chemours/DuPont). In fases van handmatige/visuele koppelingen hebben we steeds het 4-ogenprincipe toegepast en elkaars werk gecontroleerd.”

Data-analyse

Beschrijf per onderzoeksvraag:

- Welke analysemethoden hebben we gebruikt? Bijvoorbeeld regressieanalyse of multilevelanalyse.
- Hoe hebben we de uitkomsten gevalideerd? We maken hierbij onderscheid tussen interne validiteit (meten we wat we willen) en externe validiteit (in hoeverre zijn de uitkomsten generaliseerbaar). Voor interne validiteit gaat het onder meer om het geven van een onderbouwing dat het gekozen onderzoeksdesign en de methoden voor dataverzameling en analyse passen bij deze onderzoeksvragen. Bij externe validiteit gaat het bijvoorbeeld om een toelichting op de selectie van de casussen en het trekken van steekproeven.
- Hoe relateren de uitkomsten van de data-analyse aan de bevindingen en feitelijke conclusies in het rapport? Maak duidelijk welke redeneerlijn is gebruikt om te komen tot een feitelijke conclusie.

Voorbeeld: onderzoek Toegang tot langdurige zorg

We hebben eerst met logistische multilevelanalyses onderzocht in hoeverre de kans op afwijzing van een aanvraag voor langdurige zorg op individueel niveau verklaard kan worden door verschillen tussen gemeenten. Door multilevelanalyses te gebruiken houd je rekening met de mogelijke clustering van aanvragers binnen gemeenten. De kans op afwijzing van aanvragers kan namelijk niet onafhankelijk zijn omdat ze binnen een gemeente aan hetzelfde gemeentebestuur worden blootgesteld, bijvoorbeeld met betrekking tot hoe de Wmo wordt uitgevoerd. In de eerste stap van een multilevelanalyse wordt getoetst hoe sterk deze clustering is. De uitkomst van deze stap geeft gelijk antwoord op de vraag “in hoeverre hangt de kans op afwijzing van aanvragen voor langdurige zorg samen met verschillen tussen gemeenten”.

In de eerste stap hebben we een ‘leeg’ multilevelmodel geschat met kans op afwijzing als uitkomstvariabele en gemeente als level-2-variabele (maar geen andere onafhankelijke variabelen) op alle aanvragen uit gemeenten met gemiddeld meer dan 20 aanvragen per jaar. Op basis van dit model hebben we de intra class correlation (ICC) berekend. Deze correlatie geeft aan in hoeverre de variatie in waarschijnlijkheid van afwijzing tussen gemeenten groter is dan de totale variatie in kans op afwijzing. De ICC kan variëren van 0 – 1, waarbij 0 volledige onafhankelijkheid (de kans op afwijzing hangt niet van verschillen tussen gemeenten af) is en 1 volledige afhankelijkheid (binnen een gegeven gemeente is er geen verschil in kans op afwijzing).

11.3 Definities

Geef aan welke definities we hebben gehanteerd die bepalend zijn geweest in de aanpak.

Voorbeeld: onderzoek Toeslagen terugbetalen

Toeslagschulden worden in dit onderzoek gedefinieerd als de terugvorderingen die een huishouden aan de Belastingdienst verschuldigd is. Zolang er één of meer terugvorderingen lopen bij een toeslagontvanger of diens toeslagpartner, heeft hun huishouden een toeslagschuld. De terugvorderingen, per individuele toeslag en per toeslagjaar, bouwen op tot de totale toeslagschuld van een huishouden.

Hoewel *terugvorderingen* door de Belastingdienst opgelegd worden aan de toeslagaanvrager gaan we er voor dit onderzoek van uit dat in de praktijk aanvragers en partners *gezamenlijk* (als toeslaghuishouden) verantwoordelijk zijn om deze bedragen terug te betalen.

12.

Literatuur en inspiratie

12.1 International Standards of Supreme Audit Institutions

- [ISSAI 300](#) Performance audit principles
- [ISSAI 3000](#) Performance audit standard
- [GUID 3910](#) Central concepts for performance auditing
- [GUID 3920](#) The performance auditing process

12.2 Algemeen

- [Toolbox Beleidsevaluaties](#), Ministerie van Financiën
- [Performance Audit ISSAI implementation Handbook](#), INTOSAI Development Initiative (IDI)
- [Helder denken over beleidsevaluaties](#), De ArgumentenFabriek
- [Concepts of performance audit](#), European Court of Auditors

12.3 Beleidsvoorbereiding

- [Beleidskompas](#)

13.

Verklarende begrippenlijst

Activiteiten: al datgene wat in een ministerie of bij een derde wordt uitgevoerd om middelen te kunnen omzetten in prestaties voor het beleid.

Beleidsinformatie: informatie over de doelstellingen van beleid, de beoogde en/of gerealiseerde effecten, de daartoe te leveren of geleverde prestaties en de daarmee gemoeide kosten.

Beleidsresultaten: gecombineerde term voor (beleids)prestaties (output) en/of (beleids)effecten (outcome).

Bestuurlijke boodschap: het geheel van bestuurlijke uitspraken in een onderzoeksrapport: oordelen, aanbevelingen en overige kwalificaties.

Bestuurlijk oordeel: zie Oordeel.

Betrouwbaarheid: de verzamelde informatie mag geen materiële onjuistheden of onvolledigheden bevatten.

Bevinding: het resultaat van de vergelijking tussen de onderbouwde feiten over de werkelijke situatie (what is of ist) en de gestelde norm, respectievelijk de verwachte situatie (what should be of soll).

Conclusie: Een (feitelijke) conclusie is het antwoord op een onderzoeksvraag en is gebaseerd op het geheel van relevante bevindingen.

Doelgroepbereik (onderzoek naar): onderzoek naar de vraag of en hoe de uiteindelijke doelgroepen van het beleid (burgers, instellingen, bedrijven) worden bereikt.

Doelrealisatie (onderzoek naar): onderzoek naar de vraag of de met het beleid beoogde prestaties en/of effecten daadwerkelijk zijn gerealiseerd.

Doelmatigheid: kan verwijzen naar de (beleids)uitvoering, de (beleids)prestaties of de (beleids)effecten. In internationaal verband ook wel aangeduid als *efficiency* (efficiëntie).

Doelmatigheid van de (beleids)uitvoering:

- mate waarin de uitvoering van het beleid met de inzet van minder middelen gerealiseerd had kunnen worden;
- mate waarin een betere uitvoering verwezenlijkt had kunnen worden met dezelfde inzet van middelen.

Doelmatigheid van (beleids)prestaties:

- mate waarin de geleverde prestaties met de inzet van minder middelen gerealiseerd hadden kunnen worden;

mate waarin meer prestaties verwezenlijkt hadden kunnen worden met dezelfde inzet van middelen.

Doelmatigheid van (beleids)effecten (kosteneffectiviteit):

- mate waarin de gerealiseerde effecten met de inzet van minder middelen gerealiseerd hadden kunnen worden;
- mate waarin meer effecten verwezenlijkt hadden kunnen worden met dezelfde inzet van middelen.

Doeltreffendheid (onderzoek naar): onderzoek naar de vraag in hoeverre door inzet van het overheidsbeleid de gewenste beleidsresultaten worden bereikt. In internationaal verband ook wel aangeduid als *effectiveness* (effectiviteit).

Doorwerkingsonderzoek: onderzoek naar de vraag of de voorwaarden hebben gewerkt die de minister gecreëerd heeft voor een goede uitvoering van het rijksbeleid op regionaal, provinciaal of gemeentelijk niveau.

(Beleids)effecten (outcome): de gevolgen van het beleid die merkbaar zijn in de maatschappij.

Effectiviteit: zie doeltreffendheid.

Efficiëntie: zie doelmatigheid.

Externe (exogene) verklaringen: (mogelijke) verklaringen voor verschillen in doelmatigheid en doeltreffendheid die buiten de invloedssfeer van het ministerie en/of de bij het beleid betrokken organisatie liggen.

Gecontroleerde: organisatie of persoon die onderwerp van onderzoek is en waar of bij wie de Algemene Rekenkamer onderzoeksbevoegdheid heeft.

Interne (endogene) verklaringen: (mogelijke) verklaringen voor verschillen in doelmatigheid en doeltreffendheid die binnen de invloedssfeer van het ministerie en/of de bij het beleid betrokken organisatie liggen.

Kwaliteit: verschillen in de kenmerken van dezelfde prestaties of effecten.

Kosten (of lasten): aan een periode toegerekende (potentiële) uitgaven, waarbij niet het moment van betaling, maar het verbruiksmoment bepalend is.

Middelen: het ingezette personeel, het verbruikte materiaal, de ingehuurde diensten en de uitgaven aan (niet-)financiële beleidsinstrumenten (zoals subsidies, leningen en voorlichtingsprogramma's).

Neveneffecten: maatschappelijke gevolgen van beleid, anders dan de primair beoogde. Het kan hierbij gaan om bedoelde en onbedoelde neveneffecten.

Normen: standaarden of criteria, waaraan de gecontroleerde moet voldoen (what should be).

Oordeel: Resultaat van een bestuurlijke afweging naar aanleiding van de bevindingen en feitelijke conclusie (wat vinden we ervan dat de norm wel of niet is gehaald).

Prestatiegegevens: er zijn drie typen van prestatiegegevens:

- beleidsdoelstellingen, zo veel mogelijk geoperationaliseerd in termen van nagestreefde en gerealiseerde effecten, gekoppeld aan doelgroepen van beleid en een tijdshorizon;
- informatie over kostprijs en kwaliteit van de door de overheid te leveren/geleverde prestaties (producten en diensten);
- programma-uitgaven, zo veel als mogelijk zinvol onderbouwd met volume- en prijsgegevens (bijvoorbeeld het verwachte aantal begunstigden van een subsidieregeling en het bedrag per subsidie).

(Beleids)prestaties (output): de resultaten van werkprocessen die binnen een organisatie worden doorlopen om een beleidsdoelstelling te realiseren.

Relevantie: 'relevant' is die informatie die het beleid, respectievelijk de activiteiten van het betrokken ministerie of de betrokken organisatie(s) adequaat weerspiegelt.

SMART-C: norm voor de formulering van beleidsdoelen. Specifiek, meetbaar, afgestemd, realistisch, tijdgebonden en consistent.

Uitgaven: daadwerkelijk in een periode betaalde bedragen.

Vergelijkbaarheid: de verzamelde informatie dient in tijdsperspectief en tussen organisaties vergelijkbaar te zijn.

Zuinigheid: het minimaliseren van de kosten van ingezette middelen. De ingezette middelen moeten op tijd beschikbaar zijn, in gepaste hoeveelheden en gepaste kwaliteit en tegen de beste prijs. Dit valt gedeeltelijk samen met doelmatigheid van beleidsprestaties (bedrijfsvoering). In internationaal verband ook wel aangeduid als *economy*.

Algemene Rekenkamer

Postbus 20015
2500 EA Den Haag
telefoon (070) 342 44 00
voorlichting@rekenkamer.nl
www.rekenkamer.nl

Den Haag, september 2023