

Handreiking Beleidsevaluaties

Evalueren van beleid en regelgeving in de zorg- en welzijnspraktijk

Janneke Schilp
Lisanne Verweij
Manja Bomhoff
Roland Friele

<http://www.nivel.nl>
nivel@nivel.nl
Telefoon 030 2 729 700
Fax 030 2 729 729

ISBN 978-94-6122-328-9

©2015 NIVEL, Postbus 1568, 3500 BN UTRECHT

Niets uit deze uitgave mag worden veelevoudigd en/of openbaar gemaakt worden door middel van druk, fotokopie, microfilm of op welke andere wijze dan ook zonder voorafgaande schriftelijke toestemming van het NIVEL te Utrecht. Het gebruik van cijfers en/of tekst als toelichting of ondersteuning in artikelen, boeken en scripties is toegestaan, mits de bron duidelijk wordt vermeld.

Inhoud

Voorwoord	4
Inleiding.....	5
1. Van een evaluatie voorafgaand aan een voorgenomen wet tot evaluatie van een afgerond programma.....	6
2. Een verscheidenheid aan theorieën en benaderingen	10
2.1 (Quasi) experimenteel evalueren.....	10
2.2 Realistisch evalueren	12
2.2.1 Verklarende evaluatie.....	13
2.2.2 Responsieve evaluatiemodel.....	14
2.2.3 Program theory evaluation en Logic analysis.....	15
2.2.4 Participerende evaluatie	16
2.3 Evaluatiestrategie	16
3. Openen van de black box	18
3.1 Reconstructie van de beleidstheorie.....	18
3.2 Onderliggende mechanismen	20
4. Context is alles	22
4.1 De rol van context in beleidsevaluaties.....	22
4.2 Voorbeelden van context in evaluaties	23
5. Methoden voor beleidsevaluatie.....	26
5.1 Kwalitatief of kwantitatief onderzoek, of een combinatie van beide	26
5.2 Mogelijke methoden van dataverzameling	26
6. Wisselwerking tussen onderzoekers en opdrachtgevers.....	29
6.1 Wensen en verwachtingen	29
6.2 Afstemming van de vraagstelling	29
6.3 Aanpassingen tijdens beleidsevaluaties	30
6.4 Rapportage.....	31
6.5 Aanbevelingen.....	31
Nawoord.....	33
Literatuur.....	34

Voorwoord

Het NIVEL wordt regelmatig gevraagd een bijdrage te leveren aan evaluaties van beleid of programma's. De evaluatievragen van opdrachtgevers zijn zeer divers. Het kan bijvoorbeeld gaan over de vraag of een afgerond programma het beoogde doel heeft bereikt, maar het kan ook gaan over aspecten van nieuwe wetgeving die nog niet is geïmplementeerd.

Bij de beleidsevaluaties die het NIVEL uitvoert speelt de relatie tussen verschillende partijen in de zorg een belangrijke rol. Een belangrijke uitdaging is het bij elkaar brengen en interpreteren van de verschillende perspectieven van zorgaanbieders, zorggebruikers, zorgverzekeraars en de overheid. Hoewel de voorbereiding en bepaling van beleid meestal plaatsvindt op (inter)nationaal, regionaal of lokaal niveau, spelen de effecten van beleid een rol op het niveau van de dagelijkse praktijk, de zorggebruiker of de zorgverlener. Het in kaart brengen van de dagelijkse praktijk vergt meestal andere vormen van dataverzameling en analyse dan het in kaart brengen van de beleidsdoelstellingen. Onze evaluaties beslaan meestal beide domeinen en gaan in op de mogelijke relaties daartussen.

In deze handreiking geven we inzicht in de manier waarop het NIVEL de vertaalslag maakt van beleid naar de dagelijkse zorgpraktijk. Aan de hand van voorbeelden van uitgevoerde beleidsevaluaties illustreren we welke belangrijke keuzes en stappen in het evaluatieproces gezet kunnen worden en hoe dit in de praktijk is gedaan. We gaan in op een aantal theorieën en benaderingen, en reflecteren op de manier waarop deze gebruikt kunnen worden in beleidsevaluaties.

Deze handreiking is vooral bedoeld voor NIVEL onderzoekers die beleidsevaluaties uitvoeren of gaan uitvoeren en kan voor verschillende doeleinden gebruikt worden. Allereerst biedt de handreiking inzicht in belangrijke keuzes en beslissingen die in iedere evaluatie gemaakt moeten worden. De handreiking laat zien welke onderdelen in iedere evaluatie terugkomen. Daarmee geeft de handreiking handvaten om een beleidsevaluatie op te zetten en uit te voeren. Daarnaast geeft de handreiking inzicht in de ervaringen en kennis die bij het NIVEL beschikbaar zijn op het gebied van beleidsevaluaties.

Inleiding

Het NIVEL voert steeds vaker beleidsevaluaties uit. De toename in de behoefte aan goed onderzoek komt voort uit het gegeven dat organisaties zich steeds meer dienen te verantwoorden voor uitgevoerd beleid en de inzet van middelen. Beleidsevaluaties gaan na of en op welke manier en met welke middelen bepaalde gewenste effecten bereikt worden of juist niet bereikt worden. Ook wordt onderzocht of de beoogde doelgroep wordt bereikt, want de effectiviteit van een beleidsmaatregel hangt hier meestal mee samen. Als laatste geven beleidsevaluaties inzicht in de succes- en faalfactoren van beleid waarmee in de toekomst rekening kan worden gehouden voor vervolg van het beleid of in ander beleid.

Deze handreiking gaat over beleidsevaluaties. Onder de term 'beleidsevaluatie' vallen zowel evaluaties van specifieke beleidsmaatregelen als ook evaluaties van een bepaald (onderdeel van een) programma. Beleidsevaluaties kunnen op verschillende niveaus betrekking hebben. Het kan gaan om een evaluatie van een regionale beleidsmaatregel, maar bijvoorbeeld ook over een voorgenomen wet van de overheid. In het algemeen kenmerken beleidsevaluaties zich door vier verschillende fasen. Als eerste wordt de opzet van de evaluatie doordacht, wat leidt tot een evaluatiekader en onderzoeksdesign. Daarna worden op verschillende manieren gegevens verzameld om de verschillende perspectieven en effecten in kaart te brengen. Hierna vindt de analyse van alle verzamelde informatie plaats. Als laatste beoordeelt men de gevonden informatie en formuleert men conclusies van de evaluatie. Over de wijze waarop evaluaties uitgevoerd kunnen worden, zijn talloze boeken, rapporten en artikelen verschenen. In de dagelijkse praktijk worden meestal uiteenlopende onderzoeksmethoden in de verschillende fasen van een beleidsevaluatie gebruikt. Afhankelijk van het doel van de evaluatie, de precieze onderzoeksvraag, de breedte van het veld en de mogelijkheden om data te verzamelen, wordt een geschikte aanpak gekozen. Veelal in overleg met de opdrachtgever en stakeholders wordt dan de optimale mix van in te zetten methoden gekozen. De inzet van de verschillende elkaar aanvullende methoden vormt daardoor meestal de basis van evaluaties.

Leeswijzer

De hoofdstukken van deze handreiking gaan in op verschillende onderdelen en fasen van beleidsevaluaties. Sommige hoofdstukken zijn heel praktisch toepasbaar, terwijl andere hoofdstukken meer als achtergrond bedoeld zijn bij de opzet van beleidsevaluaties. In hoofdstuk 1 laten we zien welke verschillende vormen van beleidsevaluaties er zijn en hoe deze zijn toegepast in NIVEL evaluaties. In hoofdstuk 2 lichten we verschillende theorieën en benaderingen toe die over beleidsevaluaties bestaan. Hoofdstuk 3 gaat in op de voorbereidingsfase in een evaluatie, waarbij de reconstructie van de achterliggende beleidstheorie en het in kaart brengen van mechanismen aan bod komt. Hoofdstuk 4 gaat in op de diverse invloed van contextfactoren. Hoofdstuk 5 beschrijft verschillende methoden om data te verzamelen die toegepast kunnen worden in evaluaties. Hoofdstuk 6 gaat over de wisselwerking tussen opdrachtgevers en onderzoekers. De relevante onderdelen van beleidsevaluaties worden in de verschillende hoofdstukken van deze handreiking beschreven, zodat de hoofdstukken ook afzonderlijk van elkaar te lezen zijn.

1. Van een evaluatie voorafgaand aan een voorgenomen wet tot evaluatie van een afgerond programma

Er zijn verschillende vormen van beleidsevaluaties te onderscheiden. Dit onderscheid kan enerzijds worden gemaakt op basis van het moment van evalueren. Vaak gaat er een tijd overheen voordat het effect van beleid in de praktijk is terug te zien. Daarom is het moment van evalueren belangrijk. Een evaluatie kan op verschillende momenten in de beleidscyclus plaatsvinden. Beleid kan afgerond zijn en onderzocht worden op uitkomsten. Maar er kan ook gedurende de invoering van een beleidsmaatregel gekeken worden naar de opbrengsten tot dan toe. Ook kan er nog voor de start van een beleidsmaatregel gekeken worden naar de verwachte opbrengsten. Hetzelfde geldt voor programma's en wetten. Ook die kunnen voor de start, tijdens de implementatie of na afloop van de implementatie worden geëvalueerd. Het moment van evalueren wordt in literatuur veel gebruikt om onderscheid te maken in type evaluaties. *Ex ante* evaluaties vinden voor aanvang van de ingezette maatregel in de fase van beleidsvoorbereiding plaats. *Ex durante* evaluaties vinden tijdens de beleidsuitvoering plaats, en *ex post* evaluaties na afloop van de maatregel bij de beoordeling van het beleid. Met dezelfde termen kan ook een onderscheid worden gemaakt in de inhoudelijke focus van een evaluatie (De Peuter e.a., 2007), wat misschien nog van grotere relevantie is dan het onderscheid op basis van tijdsperspectief alleen. Hierbij kan het enerzijds gaan om een evaluatie van beleidsinhoud, als men het concept, structuur, instrumenten, processen en acties van het beleid onderzoekt. Anderzijds kan het gaan om een evaluatie van beleidsimpact, als men de veranderingen die beleid teweeg brengt in de maatschappelijke omgeving in kaart brengt. Op dit punt worden de termen weliswaar wat verschillend geïnterpreteerd maar kun je grofweg wel stellen dat het bij *ex ante* vragen vooral gaat over wat er zou kunnen werken en dus wat de voorwaarden zijn, ter ondersteuning van de beleidsplanning. Hierin staat voornamelijk de beleidsinhoud centraal. Bij *ex post* vragen gaat het vooral over wat er gebeurd is en wat de effecten zijn ter verantwoording van het beleid. Hierbij ligt de focus voornamelijk op de beleidsimpact. Bij *ex durante* onderzoeksvragen gaat het dan juist vooral over de procesaspecten die een rol spelen met het oog op verbetering van de uitvoering. Dit kan zowel gaan over de beleidsinhoud als de beleidsimpact.

De keuze voor een bepaald type vragen voor evaluatie kan gebaseerd zijn op verschillende overwegingen. Ieder type vragen heeft zo zijn eigen waarde en juist een combinatie van verschillende typen vragen kan het belang van een evaluatie sterk vergroten. Overwegingen die daarbij een rol kunnen spelen zijn het beschikbare materiaal of de beschikbare tijd. *Ex post* vragen, die over effecten van een wet of een beleidsprogramma gaan, vergen een dataverzameling die aan heel specifieke eisen voldoet. Goede *ex ante* en *ex durante* vragen kunnen tussentijds al inzichtelijk maken wat de sterke kanten zijn van een beleidsmaatregel of een programma en welke effecten verwacht mogen worden.

In **Box 1.1** worden drie verschillende door het NIVEL uitgevoerde evaluaties toegelicht om de diversiteit aan type vragen en evaluaties gedurende verschillende fases in de beleidscyclus inzichtelijk te maken.

Box 1.1. Voorbeelden van NIVEL evaluaties in verschillende fases in de beleidscyclus

Evaluatie van de voorgenomen invoering van de Wet cliëntenrechten Zorg (Bouwman e.a., 2012)

In 2009 besloot de Minister van VWS de gehele patiëntenwetgeving op te nemen in één brede wet voor de relatie tussen cliënten en zorgaanbieders: de Wet cliëntenrechten zorg (Wcz). Het hoofddoel van de Wcz was het versterken van de positie van patiënten of cliënten in het nieuwe zorgstelsel. Dat stelsel was tot stand gekomen met het van kracht worden van de Zorgverzekerings wet (Zvw) en de Wet marktordening gezondheidszorg (Wmg). Het wetsvoorstel Wcz van juni 2012 was een veelomvattend voorstel, o.a. bestaande uit 97 pagina's.

Het NIVEL voerde een ex ante evaluatie uit in overleg met het ministerie van VWS. Een nulmeting gaf inzicht in de toenmalige situatie vanuit het perspectief van de doelstellingen van het wetsvoorstel, voor het inwerking treden van de wet. Dit inzicht zou enerzijds gebruikt kunnen worden ter ondersteuning van het latere implementatieproces. Anderzijds zou deze nulmeting gebruikt kunnen worden als uitgangspunt om de toekomstige effecten van de wet zichtbaar te maken. Binnen de evaluatie werd ingegaan op vijf centrale thema's van de Wcz. Dit werd gedaan vanuit verschillende perspectieven: cliëntenperspectief, perspectief van de beroepsgroep van klachtenfunctionarissen, perspectief van cliëntenraden en perspectief van ziekenhuisbesturen. Reeds gepubliceerde onderzoeksgegevens werden gebruikt, net als gegevens van andere lopende onderzoeken. Ook werd een aanvullende enquête gehouden onder klachtenfunctionarissen. Tijdens het uitvoeren van de evaluatie bleek dat het onduidelijk was of het wetsvoorstel voortgezet zou worden, waarmee de directe aanleiding voor de nulmeting eigenlijk wegviel. Het uitvoeren van een evaluatie in de fase van de beleidsvoorbereiding brengt daarmee het risico met zich mee dat de resultaten onvoldoende gebruikt kunnen worden als het beleid niet uitgevoerd wordt zoals beoogd. Voordeel van een ex ante evaluatie is dat als voorafgaand aan de implementatie blijkt dat de beleidsmaatregel anders gaat werken dan verwacht of het beoogde effect niet zal kunnen bereiken, de beleidsmakers het beleid inhoudelijk nog kunnen aanpassen of eventueel voortijdig kunnen beëindigen.

Tussentijdse evaluatie van het programma Sneller Beter pijler 3 (Dückers e.a., 2006)

Het nationale programma Sneller Beter werd gestart in 2004. De derde pijler van het programma was een actieprogramma gericht op het implementeren van best practices en het borgen en verspreiden van resultaten in de curatieve zorg. Hiervoor voerden projectteams in deelnemende ziekenhuizen tegelijkertijd meerdere verbeterprojecten uit.

Het NIVEL voerde een jaar na de start van het programma Sneller Beter een ex durante evaluatie uit om de implementatie van de verbetertrajecten in de deelnemende ziekenhuizen te volgen en de effecten ervan te beschrijven. Bijzondere aandacht ging uit naar succes- en faalfactoren die een rol speelden bij implementatietrajecten in ziekenhuizen. De tussentijdse resultaten konden van waarde zijn voor de uitvoering van verbetertrajecten in ziekenhuizen, maar ook voor gelijksoortige initiatieven en programma's binnen andere zorgsectoren en groepen organisaties. Voor de evaluatie van het implementatieproces en de bereikte effecten werd gebruik gemaakt van het evaluatiemodel van Stake (Stake, 1967). Op verschillende niveaus werden gegevens verzameld: programmaniveau,

ziekenhuissectorniveau, ziekenhuisniveau en afdeling/projectteamniveau. Daarnaast werden verschillende methoden toegepast om gegevens voor de evaluatie te verzamelen: Er werden interviews afgenomen en schriftelijke vragenlijsten uitgezet onder de Raad van Bestuur/kwaliteitsfunctionarissen en projectleiders binnen ziekenhuizen. Meetgegevens op de uitkomstmaat werden aangeleverd door de deelnemende ziekenhuizen.

De bevindingen werden op de verschillende niveaus weergegeven, zodat de stakeholders de voor hen belangrijke conclusies en aanbevelingen uit het rapport konden halen. Het voordeel van deze vorm van evaluatie was dat door het tijdstip van evalueren de resultaten en aanbevelingen gebruikt konden worden als leerpunten voor het vervolg van het implementatieproces.

Maagzuurremmers: gevolgen van de veranderingen in de vergoeding (Flinterman e.a., 2014)

Sinds 1 januari 2012 worden maagzuurremmers niet meer vergoed vanuit de basisverzekering wanneer patiënten deze korter dan zes maanden gebruiken. De afschaffing van deze vergoeding kostte patiënten met een verhoogd risico op maagbloedingen eenmalig ongeveer 12 euro. Patiënten zonder hoog risico op een maagbloeding moesten een maagzuurremmer altijd al zelf betalen. De Tweede Kamer diende in april 2014 een motie in met het verzoek de effecten en neveneffecten van deze 'pakketmaatregel' te onderzoeken. De angst was dat minder hoog risicopatiënten maagzuurremmers zouden gaan gebruiken, met als negatief gevolg meer maagbloedingen.

De ex post evaluatie had als doel te onderzoeken wat het effect was van deze maatregel. Dit werd gedaan door antwoord te geven op de vragen hoeveel mensen geen maagzuurremmers gebruikten terwijl ze een verhoogd risico op maagbloedingen hadden, wat de besparing was om maagzuurremmers niet meer te vergoeden via het basispakket en wat de kosten waren van zorg voor aandoeningen die ontstaan ten gevolge van het niet gebruiken van maagzuurremmers.

Om deze vragen te kunnen beantwoorden werd gebruik gemaakt van gegevens van NIVEL Zorgregistraties eerste lijn, gedeclareerde vergoedingen op basis van Diagnose Behandel Combinaties (DBC's) en apotheekgegevens. De perioden voor en na de invoering van de maatregel werden met elkaar vergeleken.

Uit de evaluatie bleek dat patiënten met een verhoogd risico op maagbloedingen deze middelen niet minder vaak gebruikten en dat er niet meer maagzweren waren voorgekomen. Verder bleek dat met de maatregel ongeveer 25 miljoen euro per jaar bespaard zou worden.

Door het combineren van gegevens van verschillende bronnen kon onderzocht worden wat het directe effect was van de ingevoerde beleidsmaatregel op het optreden van maagzweren als uitkomstmaat.

2. Een verscheidenheid aan theorieën en benaderingen

Beleidsbeoordelingen worden opgezet en uitgevoerd vanuit bepaalde theorieën en modellen. In de afgelopen decennia is er een verscheidenheid aan theorieën en benaderingen van beoordelingen ontstaan, en zijn er heel diverse methoden die bij de uitvoering van beoordelingen toegepast kunnen worden. Deze enorme hoeveelheid aan benaderingen, opvattingen en methoden is lastig goed overzichtelijk te krijgen. Ze beschrijven verschillende belangrijke aspecten van het beoordelingsproces en werken verschillende specifieke elementen uit. In de hierop volgende paragrafen willen we inzicht geven in de verscheidenheid aan theorieën en modellen die er bestaan met betrekking tot beleidsbeoordelingen. Dit kan voor onderzoekers ter inspiratie dienen bij de opzet van beleidsbeoordelingen of als naslagwerk gebruikt worden. We beogen hiermee geen compleet overzicht van theorieën te geven, maar inzicht te geven in veelgebruikte beschikbare benaderingen bij beleidsbeoordelingen. Voor de uitvoering van beleidsbeoordelingen zijn er in de literatuur eigenlijk twee stromingen te onderscheiden: (quasi) experimenteel beoordelen en realistisch beoordelen (Boekenoogen e.a., 2009). Binnen deze stromingen zijn er verschillende theorieën en benaderingen te onderscheiden. Tussen de verschillende benaderingen zijn veel overeenkomsten te vinden en is er veel overlap in bepaalde onderdelen, maar vaak wordt de focus net anders gelegd of worden er verschillende termen gebruikt. In de praktijk worden meestal elementen van verschillende benaderingen bij de opzet en uitvoering van beoordelingen gebruikt. Ook is het mogelijk om een combinatie van zowel elementen van experimenteel beoordelen als realistisch beoordelen te gebruiken. We reflecteren op de wijze waarop de verschillende benaderingen toegepast worden in beoordelingen van het NIVEL.

2.1 (Quasi) experimenteel beoordelen

In een experimentele beoordeling vergelijkt men twee gelijke onderzoeksgroepen met elkaar. De ene onderzoeksgroep wordt blootgesteld aan een bepaalde beleidsmaatregel (beleidsinterventie), en de tweede groep niet. Verondersteld wordt dat door variaties in context uit te sluiten de beleidsmaatregel als beleidsinterventie het enige onderscheid is tussen de twee groepen. Eventuele gevonden verschillen in uitkomsten tussen de twee onderzoeksgroepen is dan alleen toe te schrijven aan de beleidsmaatregel. Echter, als men heeft vastgesteld dat een beleidsmaatregel werkt, is nog onbekend waarom het precies heeft gewerkt. In de meest zuivere vorm van experimenteel onderzoek wordt een aselecte steekproef getrokken uit de te onderzoeken populatie en wordt de onderzoeksgroep aselect in een experimentele en controlegroep verdeeld zonder dat de proefpersonen en onderzoekers weten wie in welke groep zit (dubbel blind). In de praktijk is het toepassen van deze vorm van experimenteel onderzoek binnen beleidsbeoordelingen zelden mogelijk. De inhoud en toepassing van de beleidsinterventie is niet altijd duidelijk te beschrijven. Ook kan de beleidsinterventie aangepast worden gedurende het implementatietraject. Daarnaast is het de vraag of het in de praktijk mogelijk is om twee identieke onderzoeksgroepen samen te stellen en de omstandigheden in de omgeving gelijk te houden. De invloed van omgeving kan niet uitgesloten worden van de beoordeling, omdat die veelal mee bepalen hoe een beleidsmaatregel uitwerkt op een persoon of onderzoeksgroep. Meestal is er geen expliciete controlegroep waarop de maatregel niet toegepast wordt. Soms zijn er wel mogelijkheden om een controlegroep na te bootsen, zoals

bijvoorbeeld mensen waarop een beleidsmaatregel niet van toepassing is of die er niet aan mee kunnen doen. Ook kunnen gegevens van de controlegroep geschat worden aan historische trends. Hiermee wordt de experimentele methode zoveel mogelijk benaderd. Daarom wordt dit ook wel quasi experimentele evaluatie genoemd. In de praktijk is het vaak lastig om een quasi experimentele evaluatie op te zetten. Zo kan het onethisch zijn om een bepaalde groep een beleidsmaatregel te ontzeggen. Nog belangrijker is dat het meestal organisatorisch onmogelijk is om een bepaalde groep van de gevolgen van een maatregel uit te sluiten. Ook is een experimentele opzet met een controlegroep erg kostbaar en vraagt het veel tijd.

Ondanks dat het veelal lastig is om zuiver quasi experimenteel onderzoek uit te voeren, worden er binnen het NIVEL evaluaties uitgevoerd die elementen van deze stroming toepassen. Door bijvoorbeeld twee groepen met elkaar te vergelijken waarop een verschillende beleidsmaatregel wordt toegepast of te kiezen voor een historische controlegroep te gebruiken van de periode waarin een beleidsmaatregel nog niet was ingevoerd. **Box 2.1** geeft voorbeelden van NIVEL evaluaties waarbij elementen van quasi experimenteel onderzoeken werden toegepast.

Box 2.1. Voorbeelden NIVEL quasi experimentele evaluaties

Huisarts minder gevoelig voor financiële prikkels dan gedacht (Van Dijk, 2012)

Met de zorgstelselwijziging in 2006 is het onderscheid tussen particulier en ziekenfondsverzekerden afgeschaft en werd een nieuw honoreringssysteem ingevoerd voor huisartsenzorg. Het inschrijftarief voor ziekenfondspatiënten en de betaling per contact voor particulier verzekerden werden vervangen door een gecombineerd systeem met een inschrijftarief en betalingen per contact voor iedereen.

Er werd onderzocht wat het effect was van de veranderingen in het honoreringssysteem van huisartsen en eigen bijdragen voor patiënten op de betaalbaarheid, toegankelijkheid en kwaliteit van zorg. Hiervoor werd gebruik gemaakt van longitudinale data van 2002 tot 2008 uit de database van Landelijk Informatienetwerk Huisartsenzorg (LINH). De 'difference-in-difference' benadering werd toegepast, wat inhoudt dat er zowel een interventiegroep (particulier verzekerden) en een controlegroep (ziekenfonds verzekerden) als een interventieperiode (voor 2006) en een controleperiode (na 2006) werd samengesteld.

Uit de evaluatie bleek dat bij voormalig ziekenfondsverzekerden dezelfde trends in geleverde zorg te zien waren als bij particulier verzekerden. De verandering in honorering leidde niet tot grote verschillen in het aantal contacten, de toegankelijkheid, de lengte van de contacten en het volgen van de richtlijnen. Geconcludeerd werd dat het beperkte effect van het honoreringssysteem erop wees dat de professionele moraal, 'het goede doen', de overhand had in de keuzes die huisartsen maakten in Nederland.

Het voordeel van deze benadering was dat voor zowel groep specifieke factoren (verschillen tussen particulier en ziekenfonds verzekerden) als tijd specifieke factoren gecontroleerd worden zodat het effect van verandering in honorering kon worden onderzocht.

Toename in het voorschrijven van een stoppen-met-roken-behandeling door huisartsen (Verbiest e.a., 2013)

In 2007 bracht het Nederlands Huisartsen Genootschap een richtlijn uit over stoppen met roken. Die

hield in dat huisartsen van hun patiënten gingen bijhouden of ze rookten en actief gingen adviseren over stoppen.

Onderzocht werd of er sinds de invoering van de richtlijn over stoppen met roken meer rokers waren gestopt. Hiervoor werden drie verschillende databases uit de periode 2001-2012 aan elkaar gekoppeld: gegevens over het aantal stoppers (Stivoro), gegevens over voorgeschreven hulpmiddelen om te stoppen met roken (NIVEL) en gegevens over de uitgifte van stopmedicatie door apotheken (Stichting Farmaceutische Kengetallen). De periode voor de invoering van de richtlijn (2001-2007) werd meegenomen als controleperiode en de periode na invoering van de richtlijn (2007-2012) als interventieperiode. 'Interreputed time-series' analyses werden uitgevoerd, wat inhoudt dat gekeken wordt naar een verschil in uitkomst in de perioden voor en na de implementatie van een beleidsmaatregel.

Er bleek na de invoering van de richtlijn (interventieperiode) geen significante toename te zijn van het aantal rokers dat stopten met roken. Wat wél zichtbaar effect had, was de vergoeding door de zorgverzekeraars van stoppen-met-roken-medicatie. Die richtlijn werd in 2011 ingevoerd. Als een stoppen-met-roken-behandeling vergoed werd, schreven huisartsen zo'n behandeling vaker voor. Door de indeling van de beschikbare data in een interventieperiode en een controleperiode kon een direct effect van de invoering van de beleidsmaatregel afgeleid worden. De beperking daarbij was wel dat er in de tussentijd ook andere beleidsmaatregelen waren ingevoerd die mogelijk ook een effect hadden, zoals bijvoorbeeld belastingverhoging op sigaretten en het rookverbod in publieke ruimtes.

2.2 Realistisch evalueren

Beleidsevaluaties van het NIVEL vallen meestal binnen de stroming van realistisch evalueren, omdat een experimentele opzet niet altijd mogelijk of gewenst is. Deze stroming gaat meer in op de mechanismen achter de beleidsmaatregel en verdiept zich in de context en de werking van de beleidsmaatregel. Dit realistisch evalueren werd door Pawson en Tilley in 2005 opgezet als reactie op de experimentele stroming.

Realistische evaluatie is erop gericht om bloot te leggen welke mechanismen door een bepaalde beleidsmaatregel op gang worden gebracht, in welke context dat gebeurt en wat de uitkomst daarvan is (Pawson e.a., 2005). Daarbij worden variaties in context niet uitgesloten maar juist meegenomen in de evaluatie. In de aanpak van realistische evaluatie van Pawson en Tilley (1997) dient de volgende kernvraag beantwoord te worden: Wat werkt voor wie in welke omstandigheden en waarom? Beleid zal zelf nooit rechtstreeks werken, het zijn eerder de hulpbronnen die worden verschaft aan bepaalde doelgroepen die een bepaalde gedragsverandering zullen bevorderen. Een evaluatieonderzoek moet de onderliggende mechanismen aan het daglicht brengen die door de beleidsinterventie op gang worden gebracht. Daarbij worden ook de specifieke voorwaarden geïdentificeerd die ervoor zorgen dat de onderliggende mechanismen de gewenste uitwerking heeft. Het kan natuurlijk zijn dat bepaalde mechanismen een verschillende uitwerking hebben onder verschillende omstandigheden. Aangetoond moet worden of het instrument werkt, voor wie het werkt en onder welke omstandigheden het werkt. Bij een realistische evaluatiebenadering is er specifieke aandacht voor de beleidscontext, het realiteitsgehalte van het beleid, de werkzame elementen van het beleid en de beleidsuitvoering. Deze elementen dragen bij aan het optreden en

het verklaren van de effecten. Realistische evaluatie laat ruimte voor variatie in uitkomsten over verschillende contexten heen, ongeacht of die effecten gewenst of ongewenst zijn, en bedoeld of onbedoeld. In de volgende paragrafen beschrijven we een aantal theorieën en benaderingen die vallen binnen de stroming van realistisch evalueren. Bij het NIVEL zijn talloze evaluaties gedaan die binnen de stroming van realistisch evalueren vallen of waarin aspecten van deze stroming terug komen. Box 2.2 beschrijft een voorbeeld van een uitgevoerde beleidsevaluatie die binnen de stroming realistisch evalueren geassocieerd kan worden.

Box 2.2. Beleidsevaluatie binnen stroming realistisch evalueren.

Evaluatie Gefaseerd Toezicht (Kruikemeier e.a., 2010)

In 2000 startte de Inspectie voor de Gezondheidszorg (IGZ) met de ontwikkeling en de invoering van de werkmethode Gefaseerd Toezicht. Het instrument Gefaseerd Toezicht werd ontwikkeld om een meer risico gestuurde manier van toezicht en handhaving mogelijk te maken. Met deze methode brengt de IGZ eerst met behulp van indicatoren in kaart waar de risico's voor de kwaliteit van zorg groot zijn. Vervolgens spoort zij via inspectiebezoeken of handhavingsmaatregelen zorgaanbieders aan tot kwaliteitsverbetering.

In de evaluatie van Gefaseerd Toezicht is een analyse gemaakt van de ontwikkeling van Gefaseerd Toezicht door de IGZ als geheel. Vervolgens is nagegaan hoe in de praktijk invulling is gegeven aan Gefaseerd Toezicht aan de hand van vier case-studies in verschillende zorgsectoren.

Allereerst is een reconstructie gemaakt van de beleidstheorie waarop Gefaseerd Toezicht gebaseerd was. Daarbij werd veel duidelijk over het veronderstelde werkingsmechanisme van de verschillende onderdelen van het beleid, en werden er toetsbare doelstellingen geformuleerd.

De context van Gefaseerd Toezicht speelde een belangrijke rol in de evaluatie. De periode waarop de evaluatie betrekking had bestreek een periode van 9 jaar. Rekening werd gehouden met allerlei verschillende wisselende omstandigheden en ook met het gegeven dat de effectuering van Gefaseerd Toezicht plaatsvond op verschillende niveaus. Gefaseerd Toezicht bestond uit een veelheid van instrumenten en maatregelen, die samen moesten leiden tot doelbereiking. De gevolgen van het Gefaseerd Toezicht en zijn onderdelen waren moeilijk onafhankelijk te identificeren en het beleid veranderde in de tijd. Daarom werd gefocust op de ingezette instrumenten en het beoogde effect daarvan zoals dat werd afgeleid uit de reconstructie van het beleid. Valkuil hierbij was dat onverwachte gevolgen niet gelijk zichtbaar zouden worden. Daarom werd getracht om in de onderzoeksmethode in een zekere mate van openheid te voorzien voor deze onverwachte gevolgen.

2.2.1 Verklarende evaluatie

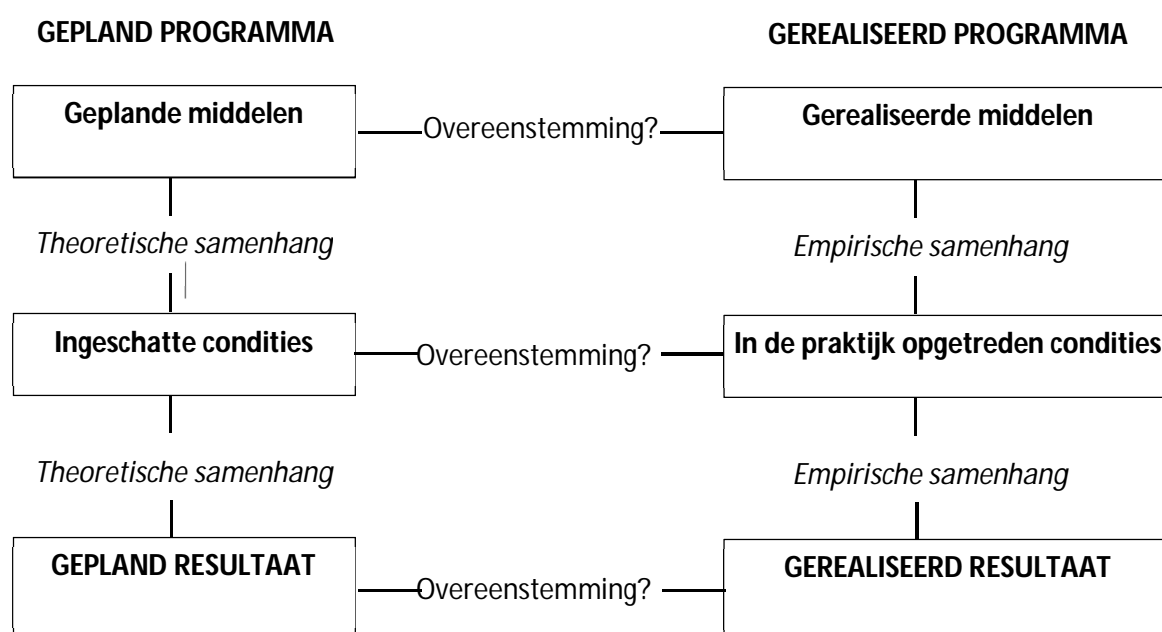
Verklarende evaluatie is een vergelijkbare benadering en heeft veel overlap met de realistische evaluatie van Pawson. (Pater e.a., 2012). Deze benadering gaat in op het 'black box probleem', waar bij beleidsevaluaties vaak sprake van is. Bij de inzet van een beleidsmaatregel worden de daaruit volgende meetbare uitkomsten gemeten, maar het is dan niet duidelijk hoe de maatregel precies heeft gewerkt en wat de verklaring is van de gevonden effecten (Pater e.a., 2012). De rol van verklarende evaluatie is het openen van de black box en het zoeken en vaststellen van verklaringen voor de verbanden tussen de beleidsinterventie en de uitkomst

(Pater e.a., 2012). Er wordt gezocht naar achterliggende mechanismen die verklaren waardoor een bepaalde beleidsmaatregel het beoogde effect heeft gehad. Of een verklaring waardoor dat juist niet (of deels) gebeurd is. De causaliteit speelt een rol bij verklarende evaluatie. Daarbij wordt aangegeven dat men pas kan spreken van causaliteit als ook de precieze werking van het mechanisme uitgelegd kan worden. Daarom gaat men op zoek naar plausibele oorzakelijke verbanden. Verklarende evaluatie is niet alleen gericht op wat beleidsinterventies doen, maar ook op hoe uitvoerders en doelgroepen handelen. Dit kan van belang zijn voor beleidsmakers, door terug te denken vanuit de uitvoerende organisaties.

2.2.2 Responsieve evaluatiemodel

Een veel gebruikte benadering bij de opzet van beleidsevaluaties en ook de evaluaties die het NIVEL uitvoert is het responsieve evaluatiemodel van Stake, waarvoor in 1967 al de basis werd gelegd (Stake, 1967). Stake ontwikkelde dit model voor toepassing in evaluaties van onderwijsprogramma's. Vervolgens werd het model toegepast in evaluaties van gezondheidszorgprogramma's. **Figuur 2.1** geeft het evaluatiemodel van Stake schematisch weer.

Figuur 2.1. Responsieve evaluatie model van Stake (1967)



Responsieve evaluatie betreft de context en relateert de verschillende perspectieven van de stakeholders aan elkaar om tot een meer geïnformeerd beeld te komen. Een uitgangspunt is dat elk perspectief slechts een deel van de werkelijkheid kan belichten. Door meerdere perspectieven en verschillende kennisvormen (ervarings-, praktijk- en wetenschappelijke kennis) in de evaluatie te betrekken ontstaat een rijker beeld van de onderzochte beleidswerkelijkheid. Een tweede uitgangspunt van 'responsieve evaluatie' is dat beleidspraktijken grillige, complexe processen zijn waarbij het doel en de middelen niet altijd eenvoudig te beschrijven zijn. Uitgevoerde acties en processen zijn daarom belangrijk om in kaart te brengen om te begrijpen waarom iets wel of niet werkt. Responsief evalueren stelt de dialoog tussen belanghebbenden centraal en brengt zo een leerproces op gang over de praktijk. Dit geeft verschillende

belanghebbenden meer inzicht en wederzijds begrip. Ook partijen die meer buiten de beleidsvormingsprocessen staan geeft het een stem en zeggenschap. En deze ervaringen en perspectieven kunnen uiteindelijk een licht werpen op de effecten, werkzaamheid en de impact van beleid.

Het responsieve evaluatiemodel bestaat uit het continu beschrijven en meten of de gestelde doelen en resultaten worden gehaald, of de geplande middelen en activiteiten ook feitelijk worden ingezet en of de verwachte succes- en faalfactoren optreden. Bij het vergelijken van het geplande met het gerealiseerde programma wordt nadrukkelijk gekeken in hoeverre de theoretisch veronderstelde relaties ook empirisch aantoonbaar zijn. De overeenstemming tussen gepland en gerealiseerd wordt tevens bepaald door condities (succes- en faalfactoren) die van invloed kunnen zijn op het implementatieproces en daarmee op de uiteindelijke resultaten. In het model van Stake kan het evalueren van het geplande programma, met de geplande middelen, ingeschatte condities en gepland resultaat, onder ex ante evaluaties worden ingedeeld. De evaluatie van het gerealiseerde programma valt onder ex-post evaluaties. Het inschatten van de mate van overeenstemming tussen het geplande en gerealiseerde programma kan men scharen onder een ex durante evaluatie.

Het model van Stake is heel geschikt voor diverse vormen van beleidsevaluaties die het NIVEL uitvoert. De evaluator zit dicht op het beleidsproces en betrokkenen participeren actief in de evaluatie, waardoor de relatie tussen verschillende partijen makkelijker gelegd kan worden (Kuindersma & Boonstra, 2005). Een beperking van dit model is dat het zich soms minder leent voor evaluaties waarin kwantitatieve uitspraken over de effecten van beleid worden verwacht.

2.2.3 Program theory evaluation en Logic analysis

De 'program theory evaluation' (PTE) gaat ook in op de black box in beleidsevaluatieonderzoek, en heeft daarom veel overlap met 'verklarende evaluatie'. De bijdrage van deze benadering is de redenering dat men verder gaat dan het beleid of programma als een black box te behandelen, maar onderzoekt waarom waargenomen veranderingen voortkomen uit beleid. Met PTE wordt nagegaan of beleid zodanig is ontworpen dat het de beoogde resultaten kan bereiken. PTE bestaat uit twee elementen: een theorie of conceptueel model over hoe het beleid de voorgenomen of geobserveerde uitkomst kan veroorzaken, en een actuele evaluatie van het beleid (Rogers e.a., 2000). Nadeel van deze benadering kan zijn dat het niet echt weergeeft hoe een beleidsmaatregel resulteert in een beoogde uitkomst, maar eerder de percepties en overtuigingen van de stakeholders over de mechanismen die tussen het beleid en de beoogde uitkomsten werken (Brousselle & Champagne, 2011). Een benadering dat voortkomt uit PTE is 'logic analysis', ofwel logische analyse. Bij een logische analyse maakt men gebruik van oordelen van experts en reviews van wetenschappelijke literatuur (Brousselle & Champagne, 2011). Logische analyse ondersteunt bij het beter begrijpen van de sterke en zwakte punten van een beleidsmaatregel en om te analyseren hoe de vormgeving van een maatregel kan bijdragen aan de verwachte resultaten. Ook kan het de kracht van de causale link tussen de beleidsmaatregel en de beoogde effecten vaststellen. Bij een logische analyse wordt eerst een logisch model gebouwd waarin de programmatheorie wordt opgesteld op basis waarvan het beleid tot het gewenste effect zou leiden. Hiervoor worden diverse databronnen gebruikt, zoals documenten en interviews met stakeholders. Daarna wordt een conceptueel framework ontwikkeld. Als laatste vindt een evaluatie van de beleidstheorie plaats. De uitwerking van de beleidsmaatregel

wordt vergeleken met de gemaakte modellen. Een participatieve benadering kan consensus creëren over wat nodig is om het beleid te verbeteren, kan de toepassing van de resultaten verbeteren en de benodigde veranderingen initiëren.

2.2.4 Participerende evaluatie

Diverse theorieën en benaderingen kunnen ondergebracht worden onder de term van 'participerende evaluatie'. Dit is empirisch onderzoek waarin betrokkenen (stakeholders) die voor- of nadeel van het onderzoek kunnen ondervinden, deelnemen als mede-onderzoekers. Deze betrokkenen zijn bijvoorbeeld opdrachtgevers, initiatiefnemers, benutters en implementeerders. Soms worden alle deelnemers, dus inclusief de professionele of primaire onderzoekers, als mede-onderzoekers aangemerkt om te benadrukken dat het in dit onderzoek gaat om samenwerking. Een belangrijk element bij dit type evaluatie is dat men niet alleen oog heeft voor aspecten die de kwaliteit van het beleid ten goede kunnen komen, maar dat men ook aandacht schenkt aan de kwaliteit van de dialoog en overleg (Van der Knaap e.a., 2012). Binnen een complexe politieke en organisatorische setting waarin vele stakeholders diverse visies hebben, is een participerende evaluatie zeker op zijn plaats. Het laat ook toe om een genuanceerder beeld te krijgen, waarin de perceptie van alle belanghebbenden aan bod komt (Verlet & Devos, 2008).

Een methode waarbij participerende evaluatie een belangrijke rol speelt, is het zogenoemde 'vierde-generatie evaluatie onderzoek' (fourth-generation evaluation), waarvan de basis ligt bij Guba en Lincoln in 1989 (Guba & Lincoln, 1989). Kenmerk van deze methode is dat personen die belang hebben bij bepaald onderzoek, een stem krijgen in de uitvoering van de evaluatie. Belanghebbenden worden niet alleen benaderd als informatieverstrekkers, maar vooral als actieve partners. Ze zijn bij voorkeur betrokken bij de formulering van de vraag en doelstelling, de selectie van betrokkenen en de interpretatie van bevindingen. Ze worden actief betrokken in de evaluatie vanuit de gedachte dat hun ervaring en deskundigheid de reikwijdte van de evaluatie vergroot, zij eerder bereid zullen zijn iets met de bevindingen te doen wanneer zij zich erin herkennen en dat zij een stem behoren te hebben omdat hun belangen worden geraakt.

Participerende evaluaties exact zoals hierboven beschreven, worden bij het NIVEL nauwelijks uitgevoerd. Wel kunnen er elementen van deze benadering toegepast worden in onze evaluaties. Zo wordt er al regelmatig gewerkt met observaties, maar worden ook verschillende vormen van co-creatie van onderzoek steeds vaker ingezet.

2.3 Evaluatiestrategie

Zoals in voorgaande paragrafen te lezen is zijn er veel overeenkomsten tussen de verschillende theorieën en benaderingen, maar zijn bepaalde elementen verschillend uitgekristalliseerd of worden verschillende termen gebruikt voor vergelijkbare concepten. Lastig is om precies aan te geven wat de verschillen zijn tussen de theorieën en benaderingen, en onder welke omstandigheden welke theorieën gebruikt zou moeten worden. Regelmatig wordt een combinatie van deze theorieën en benaderingen gebruikt als basis voor de opzet van beleidsevaluaties. Dit wordt dan echter vaak niet zo expliciet beschreven. Daarom is het lastig om specifieke voorbeelden van evaluaties te geven per theorie of benadering. Aan beleidsevaluaties van het NIVEL ligt vaak een combinatie van de het

evaluatiemodel van Stake en de theorieën van Pawson en Pater ten grondslag. Hieronder beschrijven we op basis van deze benaderingen kort de strategie die bij NIVEL beleidsevaluaties kan worden toegepast, waarbij elementen van verschillende benaderingen kunnen worden gebruikt.

Bij een beleidsevaluatie is het allereerst belangrijk om de achtergrond en de beleidstheorie in kaart te brengen die ten grondslag liggen aan de beleidsmaatregel. We beschrijven in hoofdstuk 3 hoe dit aangepakt kan worden. Als de achtergrond en de beleidstheorie in kaart zijn gebracht worden vervolgens de doelen en de geplande middelen of instrumenten van de beleidsmaatregel bepaald. Inzicht wordt gegeven in wat de bedoeling van de middelen en instrumenten was, waarvoor ze werden ingezet en wat de veronderstelde werking ervan was. Ook wordt bepaald of de instrumenten ook werkelijk in staat zouden kunnen zijn om het beoogde doel te bereiken. Van belang is om hierbij de rol van de context te betrekken, omdat de middelen in verschillende contexten anders gebruikt kunnen worden of anders kunnen uitpakken. Hier gaan in hoofdstuk 4 op in. Daarbij is het ook belangrijk om de onderliggende mechanismen die door de beleidsmaatregel op gang kunnen worden gebracht in kaart te brengen. Dit komt in hoofdstuk 3 aan de orde. Vervolgens wordt onderzocht of de instrumenten gewerkt hebben zoals ze bedacht waren en of ze het beoogde doel bereikt hebben. Mogelijk zijn er naast directe effecten ook indirecte of onverwachte effecten opgetreden door de inzet van de instrumenten. In hoofdstuk 5 worden de verschillende methoden toegelicht die kunnen ondersteunen bij het uitvoeren van een evaluatie.

3. Openen van de black box

“How to find” the program theory, when program theories not always are very explicit and there is a multiplicity of possible program theories which entails a large problem for a theory of theory-based evaluation.” (Cook, 2000), p29

Zoals in hoofdstuk 2 werd beschreven is het voor het beoordelen van het effect van een bepaalde beleidsmaatregel van belang om te zoeken naar achterliggende mechanismen die verklaren waarom de maatregel al dan niet een bepaald effect teweeg heeft gebracht. Daarbij wordt als het ware de ‘black box’ geopend en wordt gezocht naar verklaringen voor de verbanden tussen de beleidsinterventie en de uitkomst (Pater e.a., 2012). Een veelvoorkomend knelpunt waar men in beleidsevaluaties tegenaan loopt is dat de doelstellingen van een bepaalde beleidsmaatregel of programma onvoldoende helder gedefinieerd zijn. Dit maakt het lastig om na te gaan of de doelen gehaald zijn. Daarom is het bij elke beleidsevaluatie van belang dat eerst goed in kaart wordt gebracht hoe men tot de keuze voor een bepaalde beleidsmaatregel of programma is gekomen. Impliciet berust deze keuze namelijk op veronderstellingen over kenmerken en oorzaken van een geconstateerd probleem. Dit geconstateerde probleem is een verschil tussen de gewenste en de werkelijke situatie. Beleid is erop gericht om dit probleem op te heffen. Daarbij worden relaties verondersteld tussen het probleem, de inhoud van de beleidsmaatregel en de effecten. Het geheel van veronderstellingen waarop een beleidsmaatregel is gebaseerd, wordt ook wel de beleidstheorie genoemd (Boekenooen e.a., 2009). Het komt nogal eens voor dat bij invoering van een beleidsmaatregel of programma een expliciete beleidstheorie ontbreekt. In de fase van beleidsvoorbereiding is niet altijd vastgelegd welke redenering ten grondslag ligt aan de beleidsmaatregel en welk beleidsprobleem hiermee opgelost dient te worden en op welke manier. Ook kan een goede beschrijving ontbreken van de omvang van het beleidsprobleem als de beleidsmaatregel wordt geïmplementeerd. Daarom is het van belang om in beleidsevaluaties de achterliggende beleidstheorie te reconstrueren. Het expliciteren van de beleidstheorie is nodig om te beslissen over de focus en uitvoering van een evaluatie.

3.1 Reconstructie van de beleidstheorie

Een reconstructie van de beleidstheorie brengt de gedachten over hoe men tot een beleidsmaatregel is gekomen en de overwegingen daarbij in kaart. Een dergelijke reconstructie wordt als onderdeel van de logische analyse beschreven (zie paragraaf 2.2.3). Een reconstructie van de beleidstheorie is enerzijds nodig om het referentiekader in kaart te brengen, en anderzijds biedt het aangrijpingspunten voor de evaluatie. Een beleidstheorie is gebaseerd op inzichten en overtuigingen van beleidsmakers, praktijkdeskundigen en belanghebbenden. Zij hebben verschillende percepties en voorkeuren bij het ontwerpen van beleid, waarbij deze gebaseerd kunnen zijn op ervaringen, kennis, discussies, adviezen van experts, intuïtieve ideeën of aanwezige theorieën. Het gaat dus niet noodzakelijkerwijs om een evidence based theorie, maar om de theorie of gedachten die voor de beleidsmakers ten grondslag liggen aan de beleidsmaatregel.

Verschillende benaderingen kunnen gebruikt worden om de beleidstheorie te reconstrueren (Leeuw, 2003). Zo kan een beleids-wetenschappelijke benadering gebruikt worden, waarbij de evaluatie met name plaats vindt aan de hand van reviews, interviews en andere data. Ook kan een strategische benadering gebruikt worden, waarbij vier fases centraal staan: groepsindeling van stakeholders, naar boven halen van aannames in deze groepen, debat tussen de groepen en synthese. Als laatste wordt de elicitatie benadering beschreven, dat gaat over het uitlokken van reacties of het onttrekken van moeilijk verkrijgbare informatie.

Om de beleidstheorie te reconstrueren zijn verschillende methoden beschikbaar. Voorbeelden hiervan zijn: documentenanalyse, literatuuronderzoek, constructie van causale ketens en verkennende interviews met beleidsmakers. De keuze en volgorde van de toepassing van deze methoden kan verschillen tussen evaluaties, en kunnen mogelijk ook door elkaar heen lopen. Het hangt ook af van de gebruikte evaluatietheorieën en modellen welke methode men kiest. Belangrijk is om een gefundeerde overweging te maken in de keuze voor bepaalde methode(n). Daarbij is van belang aan wat voor informatie behoefte is en wat men met de resultaten wil doen. **Box 3.1** beschrijft een voorbeeld van een beleidsevaluatie waarin de reconstructie van de beleidstheorie een belangrijk onderdeel was.

Box 3.1. Reconstructie van de beleidstheorie als onderdeel van een beleidsevaluatie

Reconstructie van de beleidstheorie in een NIVEL-evaluatie van de keuze van patiënten voor zorgaanbieders (Victoor e.a., 2012)

Door invoering van het nieuwe zorgstelsel in 2006 en invoering van de marktwerking in de zorg, kregen patiënten een centrale rol. Er werd van patiënten verwacht dat zij zich als consument veel kritischer opstelden en weloverwogen keuzes maakten tussen zorgaanbieders op basis van kwaliteitsinformatie.

Om te onderzoeken wat het nieuwe zorgstelsel voor effecten had, was het van belang eerst te achterhalen wat de onderliggende assumpties waren over hoe patiëntkeuze zou moeten werken en welke impact het zou moeten hebben. Deze informatie was nodig om daarna te kunnen bepalen wat het effect van het nieuwe beleid was. Om de beleidstheorie te reconstrueren werden de beleidswetenschappelijke benadering van Leeuw (Leeuw, 2003) en verschillende methoden gebruikt. Er werd een documentenanalyse uitgevoerd met politieke documenten over patiëntkeuze. Daarnaast werden interviews afgenomen met personen die betrokken waren bij de herziening van het zorgstelsel.

Uit de reconstructie van de beleidstheorie bleek dat gedacht werd dat door het promoten van de vrije artskeuze twee belangrijke problemen opgelost zouden worden: 1) door het gecentraliseerde karakter van de gezondheidszorg werd gedacht dat de gezondheidszorg niet reageerde op patiënten en dat innovatie en flexibiliteit werden beperkt, en zouden er geen prikkels zijn voor aanbieders om hun kwaliteit en efficiëntie te verbeteren; 2) de asymmetrische patiënt-arts verhouding zou patiënten ertoe brengen om beslissingen over hun zorg aan artsen over te houden. Uit het onderzoek werd geconcludeerd dat veel aandacht was besteed aan de implementatie van diverse instrumenten om patiënten te helpen met hun keuze. Veel minder aandacht was er voor de bereidheid en mogelijkheid van patiënten om te kiezen. Uit de reconstructie van de beleidstheorie bleek dat het beleid ervan uit ging dat patiënten hiertoe bereid zouden zijn als zij de mogelijkheden

zouden hebben om te kiezen. Dit bleek in de praktijk anders te liggen. Dit toont het belang van een dergelijke reconstructie.

3.2 Onderliggende mechanismen

Bij het openen van de 'black box' in beleidsevaluaties wordt er gezocht naar onderliggende mechanismen die verklaren waardoor een bepaalde beleidsmaatregel het beoogde effect heeft gehad. Deze mechanismen kunnen beschreven worden als processen die de potentie hebben veranderingen op gang te brengen, te genereren, maar ook om veranderingen juist te voorkomen (Pater e.a., 2012). Het gaat daarbij dus niet om het veronderstelde mechanisme van stakeholders bij een bepaalde beleidsmaatregel, want dan gaat het om de beleidstheorie. Dit is een belangrijk verschil. Mechanismen zijn onderliggend en daarmee soms lastig te achterhalen (Astbury & Leeuw, 2010). Ook zijn mechanismen gevoelig voor variaties in context. Een bepaald mechanisme kan namelijk verschillend werken afhankelijk van de specifieke context. Het zijn geen universele mechanismen die overal en altijd werken. Overigens is het niet zo dat mechanismen per definitie context gebonden zijn, maar de context maakt wel uit.

Het achterhalen van de mechanismen die verklaren waarom beleid wel of niet het beoogde effect heeft gehad is enerzijds belangrijk voor het evalueren hoe en waarom het beleid werkt. Anderzijds kan deze informatie gebruikt worden voor het vormgeven en implementeren van toekomstige beleidsmaatregelen. In **Box 3.2** worden twee onderzoeken beschreven waarin het achterhalen van onderliggende mechanismen een belangrijk onderdeel was.

Box 3.2. Ontrafelen van mechanismen

Evaluatie veranderingen donorregistratiesysteem (Coppen, 2010)

In Nederland staan ongeveer 1300 mensen op de wachtlijst voor een donororgaan, maar er zijn niet meer dan zo'n 200 tot 220 donoren per jaar. Om het tekort aan organen tegen te gaan zet de overheid beleid in dat gericht is op het vergroten van het aantal post mortem donoren. In 1998 werd de Wet op de Orgaandonatie (WOD) ingevoerd.

Onderzocht werd of een verandering van het donorregistratiesysteem zou kunnen leiden tot meer donoren. Het Nederlandse beslissysteem rondom orgaandonatie werd eerst vergeleken met die van verschillende andere landen. Na correctie voor de aantallen verkeersdoden en herseninfarcten werd geconcludeerd dat een eventuele verandering van het registratie- of beslissysteem het tekort aan donoren waarschijnlijk niet zou verminderen. In de praktijk bleek namelijk dat in alle landen nabestaanden het laatste woord hebben, ongeacht het gehanteerde systeem.

In het onderzoek bleek het heel belangrijk te zijn om andere factoren die van invloed waren op de uitkomst mee te nemen, zoals structurele verschillen in relevante mortaliteit tussen landen. Omdat er sinds de invoering van de WOD tal van beleidsmaatregelen werden ingezet die mogelijk invloed hadden gehad op de donorwerving, werd deze 'black box' ontrafeld. Verschillende mechanismen bleken hierbij van belang, zoals bijvoorbeeld de ingezette voorlichtingscampagnes over orgaandonatie en donorregistratie. Deze aanvullende beleidsmaatregelen bleken nodig te zijn geweest om een stijging van post mortem orgaandonoren te realiseren.

Evaluatie van het ZonMw-programma Op Eén Lijn (De Bakker e.a., 2014)

Het ZonMw-programma Op één lijn subsidieerde tussen 2011 en 2013 in totaal 67 projecten gericht op organisatieontwikkeling en samenwerking in de eerste lijn. Deze projecten kregen daarmee een impuls om hun organisatiekracht te vergroten en de beoogde innovatie te realiseren. Uitdrukkelijk was het doel van Op Eén Lijn ook om lessen te leren uit deze projecten, door parallel hieraan een begeleidend onderzoek uit te voeren. Dit onderzoek werd uitgevoerd door het SMOEL (SamenwerkingsMonitor Op Eén Lijn) consortium, bestaande uit NIVEL (penvoerder), Jan van Es Instituut, iBMG en Panteia.

In 2010 ging het begeleidend onderzoek parallel aan wervingsrondes voor de projecten van start. In het onderzoek werd eerst inzicht gegeven in hoe succesvol de projecten binnen het programma waren op het gebied van organisatieontwikkeling voor multidisciplinaire samenwerking. Daarna werd gezocht naar welke mechanismen en factoren het succes van de projecten bepaalden. Dit werd vanuit verschillende perspectieven en met een combinatie van onderzoeksmethoden gedaan, zoals interviews, spiegelbijeenkomsten, vragenlijsten en casestudies. De resultaten van het onderzoek werden ook direct terug gekoppeld aan de betrokkenen van het project. Een sterk punt van de evaluatie was dat het rapport laat zien hoe het geïntegreerd analyseren en presenteren van resultaten tot inzichten kan leiden én meer mogelijkheden biedt om bestaande inzichten te toetsen. Er bleken verschillende factoren en mechanismen een rol te hebben gespeeld. Gekeken werd naar factoren die te maken hadden met de structuur, het proces, de uitkomsten van het project; en met de feedback tussen uitkomsten en structuur. Het rapport eindigt met een beschrijving van de mechanismen die een rol speelden bij organisatieontwikkeling en de belangrijkste lessend die daaruit getrokken kunnen worden.

4. Context is alles

Beleid vindt altijd plaats in een bepaalde context: een bestuurlijke, een sociale, een economische, een culturele maar ook een fysieke context. Op allerlei niveaus speelt context een belangrijke rol, zowel op microniveau binnen organisaties als ook op meso- en macroniveau, bijvoorbeeld binnen de politiek. Zoals Gouldner in 1955 al zei: "*context is everything*" (Gouldner, 1955). Het is van belang dat onderzoekers in alle fasen van beleidsevaluaties rekening houden met de rol van deze context, omdat het doel of de veronderstelde werking van mechanismen in verschillende contexten anders gebruikt kunnen worden of anders kunnen uitpakken.

4.1 De rol van context in beleidsevaluaties

De context speelt in alle fasen van een beleidsevaluatie en op verschillende manieren een rol. De wereld is immers continu aan veranderingen onderhevig. Dit heeft effect op de invoering en evaluatie van beleid, omdat dit niet van het ene op het andere moment gebeurt maar een langere periode beslaat. In de twee eerder besproken stromingen in beleidsevaluaties wordt verschillend met de rol van context omgegaan. In de stroming van experimenteel evalueren wordt geprobeerd de context van twee groepen zo veel mogelijk hetzelfde te houden, zodat de invloed ervan zo min mogelijk het effect beïnvloedt. In de stroming van realistisch evalueren wordt de context als onderdeel van de beleidsevaluatie beschouwd. Realistische evaluatie is erop gericht de context in kaart te brengen en te bepalen onder welke omstandigheden iets wel of niet werkt.

In de fase van beleidsvoorbereiding is het belangrijk rekening te houden met de context, omdat beleidsmaatregelen veelal ingezet worden vanuit een bepaald kader en met een bepaald doel. Door een inschatting te maken van dit kader, kan de werking van het beleid of programma beter geïnterpreteerd worden (Pater e.a., 2012).

Bij het invoeren van nieuw beleid kan het zo zijn dat de context veranderd als gevolg van het aankondigen van de invoering. Een andere mogelijkheid is dat de context al veranderd is tegen de tijd dat het beleid wordt geïmplementeerd. Bovendien kunnen andere beleidsmaatregelen mogelijk (onbedoeld) tegengestelde werkingen hebben, maar ook technologische of economische omstandigheden kunnen een rol spelen (Pawson e.a., 2005). Deze context kan de werking van de beleidsmaatregel blokkeren of juist bevorderen. Het kan ook voorkomen dat een beleidsmaatregel op een andere manier werkt dan vooraf verondersteld is. Als de context de werking blokkeert wil dat niet zeggen dat de maatregel niet deugt, maar dat de effectiviteit van de maatregel afhankelijk is van de omstandigheden waarin hij genomen wordt (Hanemaayer, 2015). De invloed van de context maakt het aantonen van causaliteit lastig. Men veronderstelt dat gevonden effecten het gevolg zijn van een beleidsmaatregel, maar het effect zou ook door externe factoren verklaard kunnen worden.

Een andere manier waarop context een rol kan spelen is bij de uitvoering van een beleidsevaluatie zelf. Beleidsevaluaties vinden namelijk plaats in natuurlijke settings, waar voorgeschiedenissen en contexten een rol spelen en waar menselijk handelen bestaat uit complexe patronen van relaties en invloeden (Pater e.a., 2012). Contextfactoren die een rol kunnen spelen bij de uitwerking van de

evaluatie zijn bijvoorbeeld: beperkte financiering, onverwachte gebeurtenissen zoals incidenten met grote media aandacht, wisseling van onderzoekers gedurende de studie of een reorganisatie bij de opdrachtgever.

4.2 Voorbeelden van context in evaluaties

In **Box 4.1** worden voorbeelden beschreven van NIVEL evaluaties waarbij de context op verschillende manieren een belangrijke rol speelde en deze met verschillende methoden in kaart werd gebracht.

Box 4.1. Rol van context in NIVEL evaluaties

Niets te klagen: onderzoek naar uitingen van ongenoegen in verzorgings- en verpleeghuizen (Bomhoff e.a., 2013)

In het onderzoek met als titel 'Niets te klagen' werd de Wet klachtrecht cliënten zorgsector (Wkcz) geëvalueerd. Dit is het klachtenbeleid in verzorgings- en verpleeghuizen. In alles fasen van de beleidsevaluatie speelde de context een rol. In dit onderzoek werd een antwoord gezocht op de vraag hoe het komt dat er zo weinig officiële klachten worden ingediend door bewoners van verpleeg- en verzorgingshuizen.

Om de klachtencontext te begrijpen werd inzicht gegeven in het bewonersperspectief, maar ook in het zorgverlenersperspectief en het beleidsperspectief. Om al deze perspectieven te leren kennen en begrijpen is een intensief, maar kleinschalig, onderzoek verricht in drie instellingen. Met behulp van observaties, semigestructureerde interviews en informele gesprekken is steeds per instelling geprobeerd een zo compleet mogelijk beeld te krijgen van de klachtencontext. De meerwaarde van een dergelijke '360-gradenblik' is dat cliënten en hun gedrag niet alleen vanuit het eigen perspectief wordt gezien maar ook in de context en het perspectief van andere relevante factoren en actoren. Voordeel is dat contextfactoren die voor bewoners zelf lastig te zien of te duiden zijn, zo ook aan het licht konden komen (Herzfeld, 2001). Bovendien gaf deze aanpak extra mogelijkheden om informatie te verkrijgen wanneer minder vanzelfsprekend uit kon worden gegaan van de uitdrukkingsvaardigheden of mondigheid van bewoners zelf en stiltes zelf ook betekenis konden hebben (Herzfeld, 2001; Gal, 2002).

Een exploratief onderzoek als 'Niets te klagen' is niet één op één reproduceerbaar. Het onderzoek werd gevormd door de gekozen instellingen, de geselecteerde respondenten, de specifieke momenten waarop het onderzoek was uitgevoerd en de onderzoekers die het onderzoek uitvoerden. Het doel van dit onderzoek was dan ook vooral om het fenomeen 'klagen' of eigenlijk 'niet-klagen' beter te begrijpen.

Monitor Zorggerelateerde Schade in Nederlandse ziekenhuizen (De Bruijne e.a., 2007; Langelaan e.a., 2010; Langelaan e.a., 2013)

Een voorbeeld van een evaluatie waar de context op een andere manier een centrale rol speelde, is de 'Monitor Zorggerelateerde Schade'. Het doel van de Monitor was om ontwikkelingen in patiëntveiligheid in Nederlandse ziekenhuizen in de tijd te volgen. De focus lag hierbij op de aard, ernst, omvang en consequenties van zorggerelateerde schade, potentieel vermijdbare schade en

potentieel vermijdbare sterfte voor patiënten in ziekenhuizen. In Nederland hebben tot nu toe drie metingen plaats gevonden om de stand van zaken te onderzoeken in 2004 (De Bruijne e.a., 2007), in 2008 (Langelaan e.a., 2010) en in 2011/2012 (Langelaan e.a., 2013).

Anders dan in het voorgaande onderzoek, richtte dit onderzoek zich niet op het verklaren van mogelijke componenten in de context die een rol spelen, maar op contextfactoren die door de tijd heen een rol spelen bij het interpreteren van resultaten uit patiëntendossiers. In het verloop van de tijd is bijvoorbeeld de patiënten mix in ziekenhuizen complexer geworden door het toenemen van de leeftijd van patiënten en het hebben van meerdere aandoeningen. De inhoud van de zorg is ook veranderd; er is sprake van meer dagbehandeling voor eenvoudige ingrepen, een kortere ligduur in het ziekenhuis en meer technische mogelijkheden. Ook werd naar aanleiding van de resultaten uit de eerste meting het landelijke VMS veiligheidsprogramma "Voorkom Schade, Werk Veilig" begin 2008 gestart. Dit programma had het doel de potentieel vermijdbare schade in vijf jaar met 50% te reduceren. Daarbij stonden tien inhoudelijke thema's centraal die alle ziekenhuizen in Nederland aangingen. De implementatie van deze thema's is door het NIVEL geëvalueerd in het laatste jaar van het Veiligheidsprogramma (De Blok e.a., 2013). Naast het veiligheidsprogramma hebben ziekenhuizen, medisch specialisten, verpleegkundigen en andere zorgverleners individueel of in groepsverband nog aanvullende verbeterinitiatieven opgezet. Alles bij elkaar werden in vijf tot tien jaar een toenemend aantal verbeterinitiatieven in de ziekenhuizen geïmplementeerd, waarmee bij de interpretatie van resultaten rekening gehouden moet worden. Desondanks bleek de kans op zorggerelateerde schade tijdens een opname in het ziekenhuis de afgelopen jaren gelijk (de kans op complicaties bij een ziekenhuisopname blijft bestaan). De potentieel vermijdbare schade door onvoldoende handelen volgens de professionele standaard of door tekortkomingen in het zorgsysteem is aanzienlijk gedaald. Dit betekent dat de patiëntveiligheid in Nederlandse ziekenhuizen aantoonbaar is verbeterd.

Evaluatie van het Programma Doelmatigheid Cardiovasculair Risicomanagement (CVRM) (Vervloet e.a., 2010)

Het Programma Doelmatigheid CVRM werd ontwikkeld door het Instituut voor Verantwoord Medicijngebruik (IVM) en het Nederlands Huisartsen Genootschap (NHG). Dit programma bevatte een breed implementatietraject voor de richtlijn 'Cardiovasculair Risicomanagement', dat erop gericht was patiënten met (een verhoogd risico op) hart- en vaatziekten te behandelen. Het doel van het Programma was het voorschrijven volgens de CVRM te bevorderen.

De evaluatie had als doel te onderzoeken of de landelijke implementatie van de richtlijn middels het programma geslaagd was. Daarnaast werd onderzocht of de systematiek van het programma werkzaam was voor de implementatie van andere richtlijnen. In de evaluatie werden verschillende methoden gebruikt, zoals vragenlijsten onder huisartsen en contactpersonen van een farmacotherapieoverleg, verzamelen van registratiegegevens over aflevermedicatie, NIVEL gegevens over voorgeschreven en afgeleverde medicatie en interviews met verschillende stakeholders. Deze evaluatie ging met name in op het proces van het programma, door te beschrijven welke beoogde middelen uiteindelijk zijn ingezet, wat het effect hiervan is geweest en welke invloed context heeft gehad. De aanbevelingen waren voornamelijk belangrijk voor de toepasbaarheid van

de gekozen opzet voor andere richtlijnen. Een resultaat van de evaluatie was dat er een aantal andere beleidsmaatregelen binnen de looptijd van het programma waren die mogelijk invloed hebben gehad op het verloop en het effect van het programma dat geëvalueerd werd. Ook waren de verschillende activiteiten en beleidsinterventies lastig afzonderlijk te evalueren, zodat geen uitspraken konden worden gedaan over welk onderdeel het meest succesvol was. Het bleek daarbij van belang de invloed van de context op het proces en de uitkomsten van beleid in kaart te brengen.

5. Methoden voor beleidsevaluatie

In dit hoofdstuk gaan we in op de verschillende beschikbare methoden die toegepast kunnen worden bij een beleidsevaluatie. Dit kan als hulpmiddel dienen bij de opzet van een beleidsevaluatie. De keuze voor een methode hangt af van verschillende factoren, zoals de aard van het onderwerp, de tijd en het budget die voor de evaluatie beschikbaar is, de beschikbare en te verzamelen gegevens maar bovenal de aard van de onderzoeksvraag. De geschiktheid van methoden verschilt ook per fase van de beleidsevaluatie. Meestal wordt de optimale mix van methoden in overleg met de opdrachtgever en andere stakeholders gekozen. Om een veelomvattend beeld te krijgen van effecten, wordt vaak gekozen voor een combinatie van meerdere methoden. Er kan een keuze gemaakt worden in het verzamelen van nieuwe gegevens, of het gebruik van al bestaande gegevens. Echte experimenten komt men in beleidsonderzoek eigenlijk niet of nauwelijks tegen. Door in evaluaties een combinatie van verschillende methoden toe te passen kan het te evalueren proces en de effecten van beleid op verschillende niveaus in kaart worden gebracht.

5.1 Kwalitatief of kwantitatief onderzoek, of een combinatie van beide

In het algemeen kan bij evaluaties een onderscheid gemaakt worden tussen kwalitatieve en kwantitatieve methoden. Met *kwantitatieve methoden* wordt nagegaan in hoeverre relaties statistisch generaliseerbaar zijn. Dit gaat over de externe validiteit van de uitkomsten. *Externe validiteit* is de mate waarin de onderzoeksresultaten generaliseerbaar zijn naar andere situaties dan die in het onderzoek. *Kwalitatieve onderzoeksmethoden* zijn meestal meer gericht op het proces, waarbij de nadruk gelegd wordt op het gedrag van stakeholders. Met een kwalitatieve methode kan onderzocht worden waarom in een bepaalde context een bepaalde activiteit al dan niet effect heeft gehad. Dit gaat over de interne validiteit van de resultaten. *Interne validiteit* is de mate waarin het redeneren binnen het onderzoek correct is uitgevoerd. In de praktijk wordt bij de gegevensverzameling van evaluaties een combinatie van zowel kwalitatieve als kwantitatieve methoden toegepast. Het gebruik van een combinatie van verschillende methoden en de interpretatie van de resultaten volgend uit de verschillende methoden wordt ook wel triangulatie genoemd. Een combinatie van kwalitatieve en kwantitatieve methoden kan een evaluatie sterker maken. Bijvoorbeeld door eerst een 'harde' inschatting van het effect van een beleidsmaatregel bloot te leggen met kwantitatieve methoden, en vervolgens met kwalitatieve methoden zoals interviews, inzicht krijgen waarom sommige instrumenten succesvol zijn en andere niet. Andersom kan het ook waardevol zijn, als we bijvoorbeeld nog niet zoveel weten over het beleid, om te beginnen met een verkennend kwalitatief onderzoek, zodat we een specifiek experiment beter kunnen richten. Het is dus geen noodzaak om te kiezen tussen methoden, maar het gaat om het kiezen van waar op welk moment de meeste behoefte aan is en wat de meeste kennis oplevert.

5.2 Mogelijke methoden van dataverzameling

Voorbeelden van methoden die in beleidsevaluaties worden toegepast zijn: interviews met sleutelfiguren, focusgroepen, deskresearch, literatuuronderzoek, casestudies, vragenlijst/enquête of

secundaire data analyse. In **Tabel 5.1** geven we een overzicht van de wijze waarop beschikbare methoden in beleidsevaluaties toegepast kunnen worden om gegevens te verzamelen.

Tabel 5.1. Toepassing van dataverzameling methoden in beleidsevaluaties

Methode	Toelichting	Toepassing
Interviews	Een onderscheid wordt gemaakt tussen (semi-) gestructureerde en niet gestructureerde (open) interviews.	De rol van interviews is afhankelijk van de fase waarin het interview wordt ingezet en de aard van informatie die gezocht wordt. In de oriëntatiefase kunnen interviews met sleutelfiguren afgenomen worden om de processen van belangrijke actoren zichtbaar te krijgen. Deze interviews zullen veelal open zijn. Ook kunnen interviews de basis bieden voor de reconstructie van de beleidstheorie of richting geven aan de evaluatie. Gestructureerde interviews bestaan uit vooraf opgestelde vragen en kunnen in een latere fase van een evaluatie gebruikt worden om te testen of de gevonden resultaten gedragen worden door stakeholders. Ook kunnen ze gebruikt worden om bepaalde effecten te verifiëren bij de beleidssubjecten.
Focusgroepen	Gestructureerde groeps gesprekken onder leiding van een ervaren gespreksleider om meningen en ervaringen over een specifiek onderwerp te inventariseren.	De kracht van de focusgroep ligt in de interactie tussen de geïnterviewden. Dit kan een belangrijke bron van informatie zijn. Met focusgroepen kunnen relevante concepten operationeel gemaakt worden. In de startfase van een evaluatie kunnen focusgroepen ingezet worden om meningen en ervaringen met het onderwerp te inventariseren en de reikwijdte te bepalen. In een latere fase van een evaluatie kunnen focusgroepen gebruikt worden om te verifiëren of de resultaten ondersteund worden.
Literatuur onderzoek	Verzamelen en analyseren van (wetenschappelijke) literatuur om zo het onderwerp in kaart te brengen.	Literatuuronderzoek kan op een niet-systematische manier in de eerste fase van een beleidsevaluatie plaatsvinden als men het onderwerp in kaart wil brengen of wil verkennen waar mogelijke hiaten liggen. De literatuur kan ook systematisch doorgenomen worden aan de hand van een systematische lijst van aandachtspunten.
Documenten analyse	Relevante documenten verzamelen en analyseren.	Om een breed beeld te krijgen van het onderwerp en de relevantie van onderdelen kan een documentanalyse worden uitgevoerd. Hiermee kunnen gegevens met verschillende invalshoeken en meningen uit verschillende bronnen gecombineerd worden. Ook maakt documentanalyse mogelijk om gegevens daterende uit verschillende tijdperioden samen te brengen. Verschillende documenten kunnen van belang zijn, zoals: beleidsstukken, rapportages, vergaderverslagen, jaarverslagen, statistieken, etc.
Observaties	Het bewust en doelgericht waarnemen van verschijnselen. Dit kan gedaan worden door het	Er zijn verschillende vormen van observaties: direct/ indirect, verborgen/ openlijk, gestructureerd/ ongestructureerd en menselijk/geautomatiseerd. Een participerende observatie is ook mogelijk, dan probeert de

	observeren van fysieke output van een activiteit, en/of bijwonen van bijeenkomsten.	onderzoeker door mee te doen met de geobserveerde personen om zo inzicht te krijgen in hoe zij dingen ervaren.
Casestudies	Diepgravend onderzoek naar enkele specifieke activiteiten. Casestudies bieden de mogelijkheid om stil te staan bij percepties van betrokkenen om geschiedenissen te reconstrueren.	Er zijn verschillende soorten casestudies. Er kan naar één specifieke case worden gekeken of naar verschillende cases tegelijk. Bronnen van casestudies kunnen bijvoorbeeld observaties, interviews, documenten, en archieven zijn. Casestudies worden meestal in een latere fase van een evaluatie uitgevoerd, zodat voldoende kennis is opgedaan om een geschikte case te selecteren.
Vragenlijst/ enquête	Schriftelijk of mondelinge vragenlijst in een steekproef van de doelgroep.	Een vragenlijst kan gebruikt worden om gegevens kwantificeerbaar te maken. Een vragenlijst kan zowel in de verkennende fase als ook in een latere fase van een evaluatie plaatsvinden. Meestal bestaat een vragenlijst grotendeels uit gesloten vragen. Open vragen kunnen de gelegenheid geven eventuele ervaringen of meningen te vermelden. Een vragenlijst kan zowel schriftelijk als digitaal verstuurd worden. Het NIVEL heeft vaste onderzoekspanels waarbij aan al dan niet een steekproef uit het panel vragenlijsten voorgelegd kunnen worden.
Database/ registraties	Gegevensverzameling bij een grote groep zorgprofessionals.	NIVEL Zorgregistraties eerste lijn is gebaseerd op longitudinale, routinematig vastgelegde gegevens in de zorg. Gegevens worden routinematig verzameld bij verschillende groepen zorgprofessionals, zoals huisartsen, fysiotherapeuten en gezondheidscentra. Koppelingen op patiëntniveau tussen verschillende registraties zijn mogelijk. Meer informatie over NIVEL Zorgregistraties eerste lijn is te vinden op de website: https://www.nivel.nl/NZR/zorgregistraties-eerstelij

6. Wisselwerking tussen onderzoekers en opdrachtgevers

Dit laatste hoofdstuk gaat in op de interactie tussen onderzoekers en hun opdrachtgevers op verschillende onderdelen en in verschillende fasen van beleidsevaluaties. Gedurende een beleidsevaluatie zal continu afstemming moeten plaatsvinden met de opdrachtgever en andere stakeholders. Heldere afstemming is belangrijk om uiteindelijk tot resultaten te komen die antwoord geven op de juiste vragen en ook tot aanbevelingen te komen die bruikbaar zijn in de reflectie op het beleid.

6.1 Wensen en verwachtingen

Bij de start van een evaluatie worden duidelijke afspraken gemaakt over de wensen en verwachtingen van beide partijen. De opdrachtgever spreekt zijn wensen uit ten aanzien van de producten, zoals een rapport, presentatie, artikel of conferentie. Ook de wensen van de opdrachtgever met betrekking tot de hardheid van de conclusies moet afgestemd worden met de onderzoekers. Nagedacht moet worden over hoe hard conclusies getrokken kunnen worden, gezien de aard van de vraag en de stand van zaken in het veld. Mogelijk kunnen resultaten gevoelig liggen als het om politiek beladen onderwerpen gaat en moeten de onderzoekers om het onderzoek goed uit te kunnen voeren kennis van deze gevoeligheid opdoen. Soms zullen de verwachtingen van de opdrachtgever bijgesteld moeten worden, omdat het bij veel onderwerpen niet mogelijk is met hele harde conclusies te komen. In overleg met de opdrachtgever wordt ook een realistische planning afgesproken. Door de stappen van de beleidsevaluatie van tevoren gedetailleerd uit te werken in een plan van aanpak kan er voldoende tijd voor de verschillende onderdelen ingepland worden. Praktische afspraken, zoals wat de eindproducten zijn en wanneer naar buiten getreden mag worden, en de planning van terugkoppelmomenten, kunnen hierin ook opgenomen worden. De opdrachtgever dient in te stemmen met het plan van aanpak, zodat hier later op terug gevallen kan worden. Vooral bij onderzoek waarbij de resultaten mogelijk politiek gevoelig zijn, is dit belangrijk zodat van tevoren duidelijk is wat de opdrachtgever kan verwachten. Het kan zo zijn dat als methoden in het plan van aanpak niet goed afgestemd zijn met de opdrachtgever of stakeholders, methoden achteraf bekritiseerd worden. Om dit voor te zijn kan maar beter in een vroeg stadium draagvlak voor de methoden zijn georganiseerd.

6.2 Afstemming van de vraagstelling

Bij het ontwerp van de evaluatie is het van belang dat onderzoekers en opdrachtgevers de vraagstelling goed met elkaar afstemmen. De aanleiding van de onderzoeksvraag moet helder zijn, waarbij het van belang kan zijn om eerst duidelijk in kaart te brengen welk probleem ten grondslag ligt aan de vraag. Vragen die hierbij een rol kunnen spelen zijn wat men precies wil weten, welke context van belang is, en over welke populatie determinanten, beleidsinterventie en uitkomsten het precies gaat. En welke relaties hiertussen van belang kunnen zijn.

Het formuleren van een volledige en juiste doelstelling en de daaraan gekoppelde onderzoeksvragen is een cruciale fase in een beleidsevaluatie. Om te beoordelen of een bepaalde beleidsmaatregel het beoogde effect heeft gehad, is het als eerste onderdeel in een beleidsevaluatie van belang om te bepalen welk effect men met het beleid beoogde te bereiken. Dit kan in het plan van aanpak worden

opgenomen als de fase waarin het evaluatief kader wordt opgesteld. Als de precieze doelstellingen of evaluatievragen van een beleidsmaatregel niet helder omschreven zijn of zelfs tegenstrijdig blijken, is het lastig om na te gaan of de doelen gehaald zijn. Een doelstelling kan op het eerste gezicht best duidelijk zijn, maar toch niet specifiek genoeg zijn voor een evaluatie. Als bijvoorbeeld de doelstelling was om een maatregel in te voeren om de kwaliteit van zorg te verbeteren, kan onduidelijk zijn wat men precies verstaat onder kwaliteit van zorg, en wanneer er sprake is van een verbetering in kwaliteit. Wat kan helpen is om de doelstellingen van een evaluatie zo te formuleren dat deze Specifiek, Meetbaar, Acceptabel, Realistisch en Tijdgebonden, oftewel SMART, zijn. De kunst is om goede heldere vragen te stellen en de vraag voldoende in te perken zodat deze hanteerbaar blijft. Vervolgens formuleert men voor deze doelstellingen overeenstemmende indicatoren aan de hand waarvan de mate van doelbereiking gemeten kan worden. Doelstellingen kunnen gaandeweg een project ook wijzigen. Bovendien kunnen beleidsmakers de doelstellingen van beleid achteraf ook anders formuleren dan vooraf.

Het type onderzoeksvragen verschilt per fase van de beleidscyclus (Bensing e.a., 2003). In de fase van *beleidsvoorbereiding* beogen de onderzoeksvragen vooral de grootte van het probleem en de mogelijke consequenties te verkennen. De belangrijkste vraag daarbij is meestal: is er (echt) een probleem dat aangepakt dient te worden?

In de fase van *beleidsbepaling* richten onderzoeksvragen zich op de oorzaken van het probleem, om deze informatie te vertalen naar aanbevelingen voor het beleid. De hoofdvraag daarbij is meestal: wat zijn de oorzaken van het probleem en wat is de haalbaarheid en uitvoerbaarheid van de plausibele oplossingen?

In de fase van *beleidsbeoordeling* richten onderzoeksvragen zich op het proces en de resultaten van beleid. Meestal is de hoofdvraag: werkt de beleidsmaatregel zoals verwacht, en wat zijn de verwachte en onverwachte effecten? Belangrijk is om uiteindelijke conclusies van dit proces ter validatie voor te leggen aan de opdrachtgever of beleidsmakers zelf.

Belangrijk is om bij het ontwerp van de evaluatie na te denken over wat voor soort antwoorden de opdrachtgever zoekt, wie de mogelijke gebruikers van de resultaten zijn en hoe deze benaderd kunnen worden vanaf de start van de evaluatie (Bensing e.a., 2003). Naast de opdrachtgever, kunnen verschillende belanghebbenden wisselende motieven, opvattingen en belangen in de evaluatie hebben. Deze zullen in kaart moeten worden gebracht om de focus van de evaluatie te kunnen bepalen.

Naast de inhoudelijke afstemming van de vraagstelling is de financiële afstemming met de opdrachtgever bij een beleidsevaluatie natuurlijk ook heel belangrijk. Daarbij is het goed om te realiseren dit altijd gaat in overleg met de programmaleider en dat daarnaast ook een beroep gedaan kan worden op de kennis en ervaring van de adjunct-directeur bedrijfsvoering.

6.3 Aanpassingen tijdens beleidsevaluaties

Het kan voorkomen dat gedurende de evaluatie wensen en verwachtingen of de doelstelling van de opdrachtgever veranderen. Dit kan bijvoorbeeld komen door invloeden van het te evalueren beleid of programma, maar ook door nieuwe inzichten die gaandeweg ontstaan uit contextfactoren. Daarom is het belangrijk om ook gaandeweg de evaluatie de doelstelling en evaluatievragen steeds goed met elkaar af te stemmen zodat de juiste vraagstellingen worden uitgewerkt. Een open communicatie tussen de verschillende partijen gedurende beleidsevaluaties bevordert betrokkenheid en draagt bij aan openheid van opdrachtgevers ten opzichte van het onderzoek en van onderzoekers ten opzichte van de problemen in de praktijk van het beleid en de politiek. De

begrenzing en beperkingen van een evaluatie moeten duidelijk geschetst worden voor opdrachtgevers en belanghebbenden om discrepantie achteraf te voorkomen. Aanpassingen in de evaluatie leveren al gauw extra werk op. Een goede afstemming tussen opdrachtgever en de onderzoeker over de inhoud van dit extra werk en de oplossing ervan moet duidelijk gecommuniceerd worden. Een extra doelstelling of evaluatievraag kan invloed hebben op de dataverzameling als er meer gegevens nodig zijn, maar kan bijvoorbeeld ook extra werk betekenen in de analysefase of de rapportagefase. Dit extra werk brengt dan vaak ook extra kosten met zich mee.

6.4 Rapportage

Wat precies het eindproduct van een evaluatie is, hangt af van het specifieke doel van de evaluatie en de context. Over welke keuzes hierbij gemaakt moeten worden en hoe dit proces het beste aangepakt kan worden gaat de overzichtsstudie over het nieuwe stijl NIVEL rapport (2015). De rapportage kan een opsomming geven van resultaten of een aantal boodschappen geven waar opdrachtgevers zelf hun voordeel mee kunnen doen. Ook kan de rapportage gerichte aanbevelingen bevatten. Verantwoording van de resultaten van onderzoek is belangrijk. Dit kan lastig zijn bij beleidsevaluaties, omdat een evaluatie op verschillende manieren impact kan hebben. Het gaat niet alleen om wetenschappelijke impact, maar ook juist om de maatschappelijk impact van de resultaten (Bensing e.a., 2003). Een belangrijk doel van een beleidsevaluatie is dat de gegenereerde informatie uiteindelijk wordt benut en de kwaliteit van beleid helpt verbeteren. Daarvoor is het belangrijk dat de resultaten een informatieve waarde hebben, herkenbaar en toepasbaar zijn en dat een inschatting wordt gemaakt van de mate van bedreiging: de mate waarin resultaten beleidsvoerders, politici of andere betrokkenen in een lastig parket kunnen brengen (Ministerie van Financiën, 2003). Als de inschatting is dat die mate van bedreiging groot zal zijn moet men extra aandacht besteden aan het bepalen van het moment en de wijze van openbaarmaking van de resultaten. Bij de rapportage van onderzoeksresultaten is er meestal een vertaling naar de doelgroepen nodig, wat ook wel disseminatie wordt genoemd. De rapportage van de resultaten moet namelijk toegankelijk zijn voor potentiële lezers. Mogelijke knelpunten bij het openbaar maken van resultaten kunnen zijn dat er sprake is van geringe medewerking van de te evalueren beleidsinstantie of uitvoerder omdat men eigenlijk niet wil veranderen, dat de uitkomsten van de evaluatie niet bevallen, of dat de opdrachtgever vanwege politieke redenen slechts een beperkte evaluatie wil waardoor geen goed totaal oordeel gegeven kan worden (Boekenoogen e.a., 2009). Om deze knelpunten te voorkomen kan het helpen om vooraf goed met de opdrachtgever af te stemmen wat de verwachtingen zijn, en wat er met de resultaten gedaan zal worden. Daarbij is niet openbaarmaking overigens geen optie. Het NIVEL is statutair verplicht om alle onderzoeksresultaten en conclusies van onderzoek openbaar te maken, teneinde de inhoud en de methodiek van het uit te voeren onderzoeksproject te kunnen laten beoordelen door het wetenschappelijk forum en betrokkenen in de samenleving van de resultaten te laten kennisnemen.

6.5 Aanbevelingen

Als er aanbevelingen worden gedaan in een evaluatierapportage, is het meestal gewenst dat deze zo concreet mogelijk geformuleerd en technisch uitvoerbaar zijn. Belangrijk is dat aanbevelingen voortkomen uit de resultaten van het onderzoek. Vaak dient in de aanbevelingen rekening gehouden te worden met de haalbaarheid op bijvoorbeeld politiek, maatschappelijk en financieel oogpunt.

Belangrijk is ook om helder te zijn over aan wie welke aanbeveling gericht is. De evaluaties van het NIVEL zijn er niet op gericht specifieke beleidsdoelen te formuleren, maar in de aanbevelingen kunnen wel eventuele tegenstrijdigheden in beleid worden opgenomen of voorwaarden benoemd worden om bepaalde effecten te kunnen realiseren.

Bij het opnemen van aanbevelingen in een rapportage is het belangrijk rekening te houden met het feit dat beleid niet tijdloos is. Bij het te direct overnemen van of refereren aan beleidstermen of doelen kan het risico ontstaan dat de onderzoekers te veel worden gezien als een verlengstuk van de opdrachtgever. Wanneer het politieke klimaat dan enige tijd later is veranderd, kan het zijn dat de onderzoekers alsnog worden geassocieerd met het uit de gratie geraakte beleidsdoel. Een manier om dit risico te vermijden is door vooral conditionele aanbevelingen te formuleren, zoals bijvoorbeeld: "Als het beleid gericht is op... dan lijkt ... hiervoor een belangrijke conditie".

Afhankelijk van het doel van een evaluatie, kan het belangrijk zijn om de beoogde gebruikers van de resultaten te betrekken bij het formuleren van conclusies en aanbevelingen. Door mensen te betrekken die goed ingebed zijn in de omgeving waar de evaluatie betrekking op heeft kunnen resultaten en aanbevelingen beter aansluiten op de doelgroep. Dit kan gedaan worden door bijvoorbeeld een bijeenkomst te organiseren met stakeholders, en daar uiteindelijke oplossingen, boodschappen, conclusies en aanbevelingen op basis van de voorlopige resultaten te bespreken. Hiermee kan tegelijkertijd draagvlak worden gecreëerd bij de doelgroep.

Nawoord

Met deze handreiking hebben we geprobeerd onderzoekers inzicht te geven in belangrijke onderdelen van beleidsevaluaties, en keuzes en beslissingen die daarbij een rol spelen. We hebben hiervoor gebruik gemaakt van een aantal voorbeelden van evaluaties die zijn uitgevoerd bij het NIVEL om te laten zien hoe verschillende aspecten hierin terugkomen. Hierbij gaat het om een kleine selectie van voorbeelden uit een enorme verzameling van evaluaties die binnen het NIVEL zijn uitgevoerd. Natuurlijk zijn er nog veel meer voorbeelden van beleidsevaluaties te vinden binnen de NIVEL publicaties. Voor de meest gebruikte (internationale) literatuur zijn we op zoek gegaan naar publicaties over beleidsevaluaties op Pubmed en google scholar. Daarnaast zochten we in de referentielijst van reeds gevonden literatuur. We merken hierbij op dat we geen compleet beeld geven van de beschikbare literatuur over dit onderwerp, maar dat we geprobeerd hebben een zo relevant en duidelijk mogelijk overzicht te geven voor onderzoekers. Daarbij hebben we beoogd te achterhalen of er nieuwe theorieën of modellen zijn ontwikkeld die gebruikt zouden kunnen worden in evaluaties. Zoals te lezen is in de handreiking zijn er verschillende theorieën en modellen gevonden die nieuwe elementen beschrijven of aspecten op een andere manier benaderen. De totale lijst met gevonden literatuur is verzameld in een Endnote bestand en is op een openbare plaats via intranet te vinden.

We hopen dat deze handreiking onderzoekers handvaten geeft voor het opzetten en uitvoeren van beleidsevaluaties.

Literatuur

- Astbury B, Leeuw FL. *Unpacking black boxes: mechanisms and theory building in evaluation*. American Journal of Evaluation, 2010; 31(3): p.363-381
- Bensing JM, Caris-Verhallen WM, Dekker J, Delnoij DM, Groenewegen PP. *Doing the right thing and doing it right: toward a framework for assessing the policy relevance of health services research*. Int J Technol Assess Health Care, 2003; 19(4): p.604-612
- Boekenoogen M, van Hoesel P, Prince Y, Verheijen C. *Methoden van beleidsonderzoekers: creatief en oplossingsgericht*. Den Haag: Uitgeverij LEMMA, 2009
- Bomhoff M, Paus N, Friele RD. *Niets te klagen: Onderzoek naar uitingen van ongenoegen in verzorgings- en verpleeghuizen*. Utrecht: 2013
- Bouwman R, Bomhoff M, Friele RD. *Kennisvraag: Vijf patiëntenrechten uit het wetsvoorstel Wet cliëntenrechten zorg (Wcz) anno 2012*. 2012
- Brousselle A, Champagne F. *Program theory evaluation: Logic analysis*. Eval Program Plann, 2011; 34(1): p.69-78
- Cook TD (2000). The false choice between theory-based evaluation and experimentation. Program theory in evaluation: Challenges and opportunities. New directions in evaluation. San Francisco, Jossey-Bass. **87**: 27-35
- Coppen R. *Organ donation, policy and legislation: with special reference to the Dutch Organ Donation Act (proefschrift)*. Utrecht: 2010
- De Bakker D, Batenburg R, De Bont A, Bruijnzeels M, Maaijen M, Poortvliet P, e.a. *Georganiseerd op weg in de eerste lijn - Resultaten van SMOEL - Eindrapport*. 2014
- De Blok C, Koster E, Schilp J, Wagner C. *Implementatie VMS Veiligheidsprogramma. Evaluatieonderzoek in Nederlandse ziekenhuizen*. Utrecht/Amsterdam: NIVEL/EMGO+, 2013
- De Bruijne MC, Zegers M, Hoonhout LHF, Wagner C. *Onbedoelde schade in Nederlandse ziekenhuizen: dossieronderzoek van ziekenhuisopnames in 2004*. Amsterdam: I. v. E. G. Onderzoek, 2007
- De Peuter B, De Smedt J, Bouckaert G. *Handleiding beleidsevaluatie deel 1: evaluatiedesign en -management*. 2007
- Dückers M, De Bruijn M, Wagner C. *Evaluatie Sneller Beter pijler 3, De implementatie van verbeterprojecten in de eerste acht ziekenhuizen*. Utrecht: 2006
- Flinterman L, Hek K, Korevaar J, van Dijk L. *Maagzuurremmers: gevolgen van de veranderingen in de vergoeding*. Utrecht: 2014
- Gal S (2002). Between speech and silence. The anthropology of politics; a reader in ethnography, theory and critique. Oxford, Blackwell publishing

- Gouldner A. *Wildcat strike: a study in workmanagement relationships*. London: Routledge, 1955
- Guba E, Lincoln Y. *Fourth Generation Evaluation*. Beverly Hills, 1989
- Hanemaayer D (2015). Van <http://www.beleidsevaluatie.info>
- Herzfeld M. *Anthropology: theoretical practice in culture and society*. Oxford: Blackwell Publishers, 2001
- Kruikemeier S, Schäfer WLA, Coppens R, Friele RD. *Evaluatie Gefaseerd Toezicht*. Utrecht: 2010
- Kuindersma W, Boonstra F. *Methoden van beleidsevaluatie onder de loep; een zoektocht naar nieuwe vormen van beleidsevaluatie voor het Milieu- en Natuurplanbureau*. 2005
- Langelaan M, Baines RJ, Broekens MA, Siemerink KM, van de Steeg L, Asscheman H, e.a. *Monitor Zorggerelateerde Schade 2008*. Utrecht: Nivel, 2010
- Langelaan M, De Bruijne MC, Baines RJ, Broekens MA, Hammink K, Schilp J, e.a. *Monitor Zorggerelateerde Schade 2011/2012: dossieronderzoek in Nederlandse ziekenhuizen (in Dutch)*. Amsterdam/ Utrecht: 2013
- Leeuw FL. *Reconstructing program theories: methods available and problems to be solved*. American Journal of Evaluation, 2003; 24(1): p.5-20
- Ministerie van Financiën. *Handreiking Evaluatieonderzoek ex post*. Den Haag: 2003
- Pater C, Sligte H, van Eck E. *Verklarende evaluatie: een methodiek*. Amsterdam: 2012
- Pawson R, Greenhalgh T, Harvey G, Walshe K. *Realist review--a new method of systematic review designed for complex policy interventions*. J Health Serv Res Policy, 2005; 10 Suppl 1: p.21-34
- Rogers PJ, Petrosino A, Huebner TA, Hacsí TA. *Program theory evaluation: Practice, promise and problems. New directions for evaluation*. San Francisco: Jossey-Bass, 2000
- Stake RE. *The countenance of educational evaluation*. Teacher College Record, 1967; 68(7): p.523-540
- Van der Knaap P, Gorree M, Israel E. *Subsidie-evaluaties onder de loep: onvoldoende evidence voor effectiviteit*. Bestuurskunde, 2012; 2: p.30-40
- Van Dijk C. *Changing the GP payment system: do financial incentives matter? (proefschrift)*: NIVEL, 2012
- Verbiest ME, Chavannes NH, Crone MR, Nielen MM, Segaar D, Korevaar JC, e.a. *An increase in primary care prescriptions of stop-smoking medication as a result of health insurance coverage in the Netherlands: population based study*. Addiction, 2013; 108(12): p.2183-2192
- Verlet D, Devos C. *Over beleidsevaluatie: van theorie naar praktijk en terug*, 2008
- Vervloet M, Brabers AEM, Pisters MF, van Dijk L. *Evaluatie van het Programma Doelmatigheid Cardiovasculair Risicomanagement (CVRM)*. Utrecht: 2010

Victoor A, Friele RD, Delnoij DM, Rademakers JJ. *Free choice of healthcare providers in the Netherlands is both a goal in itself and a precondition: modelling the policy assumptions underlying the promotion of patient choice through documentary analysis and interviews*. BMC Health Serv Res, 2012; 12: p.441