

**Ex post evaluatie
verkeersveiligheidsmaatregelen**

Datum	14 maart 2011
Status	definitief

Ex post evaluatie verkeersveiligheidsmaatregelen

Datum 14 maart 2011
Status definitief

Colofon

Titel	Ex post evaluatie verkeersveiligheidsmaatregelen
Uitgegeven door	Dienst Verkeer en Scheepvaart
Informatie	DVS loket Telefoon 088 7982555 Email: dvsloket@rws.nl
	Contactpersoon DVS: drs. W. Vermeulen
Uitgevoerd door	DVS, SWOV en KiM
Datum	14 maart 2011
Copyright	Rijkswaterstaat Dienst Verkeer en Scheepvaart Delft 2011

Disclaimer

De Dienst Verkeer en Scheepvaart (DVS), en degenen die aan deze publicatie hebben meegewerkt, hebben de in deze publicatie opgenomen gegevens zorgvuldig verzameld naar de laatste stand van wetenschap en techniek. Desondanks kunnen er onjuistheden in deze publicatie voorkomen. Rijkswaterstaat sluit iedere aansprakelijkheid uit voor schade die uit het gebruik van de hierin opgenomen gegevens mocht voortvloeien.

Inhoud

Voorwoord	blz.	6
1. Samenvatting		7
2. Inleiding		9
2.1 Aanleiding voor dit onderzoek		9
2.2 Stand van zaken		9
2.3 Doel van het onderzoek		10
2.4 Onderzoeksvragen		10
3. Begripsbepaling		11
3.1 Evaluaties in fasen van het beleidsproces		11
3.2 Beleidsevaluatie		11
3.3 Benaderingswijzen		13
3.4 Ex ante beleidsevaluatie		14
3.5 Ex durante beleidsevaluatie		16
3.6 Ex post beleidsevaluatie		16
3.7 Beschouwing		17
4. Doelen en betekenis ex post beleidsevaluatie		19
4.1 Verantwoorden		19
4.2 Leren of kennisgeneratie		20
4.3 Proces, prestatie en product		20
4.4 Betekenis ex post beleidsevaluatie		21
5. Methodologie ex post evaluatie		23
5.1 Inleiding		23
5.2 De ideale opzet – experimenteel onderzoek		23
5.3 De praktijk – observationeel onderzoek		24
5.4 Conclusies		26
6. Beschouwing: wat is zinvol en haalbaar?		28
6.1 Complexiteit van de problematiek		28
6.2 Afwegingen		29
6.3 Zinvol ex post evalueren		30
7. Conclusies		32
7.1 Conclusies		32
Literatuur		35
Bijlage 1: Methodologische uitwerkingen		39
Bijlage 2: Beleidsdocumenten ex post vragen		65

Voorwoord

Dit rapport is geschreven naar aanleiding van vragen in het Algemeen Overleg (AO) van de Tweede Kamer over de effectiviteit van maatregelen die beogen de verkeersveiligheid te verbeteren. De effectiviteit wordt vaak ex ante (vooraf) vastgesteld, om de beleidsmaatregel mede te rechtvaardigen. Of de inschatting wordt gerealiseerd wordt vaak niet ex post (naderhand) vastgesteld. In dit rapport wordt verkend welke mogelijkheden en beperkingen er zijn voor een evaluatie ex post.

Het rapport is tot stand gekomen op grond van de inbreng van Rijkswaterstaat, Dienst Verkeer en Scheepvaart (DVS), de Stichting Wetenschappelijk Onderzoek Verkeersveiligheid (SWOV) en het Kennisinstituut voor Mobiliteitsbeleid (KiM). Van de SWOV is Jaques Commandeur integraal verantwoordelijk voor de inhoud van bijlage 1, Methodologische uitwerkingen. Rijkswaterstaat is als eindredacteur de SWOV en het KiM erkentelijk voor hun bijdragen en adviezen.

1. Samenvatting

Dit rapport, dat tot stand is gekomen door samenwerking van RWS-DVS, SWOV en KiM bevat het resultaat van een literatuurstudie naar de mogelijkheden van ex post evaluatie van verkeersveiligheidsmaatregelen.

Om deze mogelijkheden in kaart te brengen is het allereerst van belang de plaats van ex post evaluatie in de context van beleidsevaluatie te schetsen. Het overheidsbeleid moet kunnen worden verantwoord. De verantwoording betreft zowel de evaluatie van de doelen van het beleid als de effecten van het gevoerde beleid. Elementen zoals het bereik van doelgroepen, de wijze waarop en de mate waarin beleid tot stand is gekomen, de efficiëntie van beleid en de effectiviteit maken daar deel van uit.

Daarnaast blijkt het leren op grond van evaluatie waardevol te zijn, om bij analoge beleidsprocessen betere of efficiëntere besluiten te nemen, of om vervolgevaluaties beter op te zetten en uit te voeren.

Bij het opzetten van een ex post beleidsevaluatie moeten veel vragen worden beantwoord. Belangrijke vragen betreffen het doel van de evaluatie, de mogelijkheden om een betrouwbare en valide evaluatie te ontwerpen, de omvang en zwaarte van de evaluatie, de meest gewenste uitvoeringsperiodes, de verwachte resultaten en de beleidsruimte om aan de resultaten besluiten te koppelen en de kosten van uitvoering. Om deze vragen te beantwoorden kan een evaluatieprogramma worden opgesteld, waarin voornoemde elementen zijn opgenomen. Hierbij is een keuze tussen verschillende benaderingswijzen van evalueren mogelijk, van puur het meten van effecten tot het in samenspraak overdenken en verklaren van de genomen maatregelen in de praktijk. Beleidsevaluatie richt zich op de doelen van het beleid (zijn deze bereikt), de processen en de wijze van uitvoering van beleid in de procesevaluatie, de werkingsfactoren die tot het al dan niet bereiken van de doelen hebben bijgedragen, de doelmatigheid van beleid en de doelmatigheid van de beleidsuitvoering.

Voorts wordt ingegaan op methodologische vraagstukken; dit betreffen de vragen welke methoden gebruikt kunnen worden, en of voor de toepassing van de methoden de geschikte data voorhanden zijn. Hier speelt de causaliteitsvraag: kunnen gevonden resultaten inderdaad aan de ingevoerde maatregel(en) worden toegeschreven, of spelen andere factoren een rol?

Idealiter kan deze vraag worden beantwoord door het opzetten van een zuiver experiment, waarbij variabelen eenduidig worden benoemd en gemeten. Vervolgens worden de effecten van het al dan niet in werking zijn van de maatregel vastgesteld door aselect een steekproef van voldoende omvang te kiezen van situaties waarin de maatregel wel (experimentele groep), en waarin deze niet wordt toegepast (de controlegroep), met een voor- en een nameting.

Deze ideale situatie blijkt in het verkeer zelden of nooit te kunnen worden gerealiseerd. Dan zijn er alternatieve opzetten, zoals een experiment zonder voormeting of een experiment zonder controlegroep, maar dan kunnen doorgaans geen harde conclusies over de oorzaak van veranderingen worden getrokken. Alternatieve verklaringen voor een eventueel gevonden effect van een maatregel kunnen zijn een algemeen dalende trend, regressie naar het gemiddelde, verandering in de registratie, effecten van andere gebeurtenissen dan de maatregel zelf, of een combinatie van deze.

Als het niet mogelijk is om random te selecteren of toe te wijzen aan experimentele en controlegroep, dan wordt gekozen voor observationeel onderzoek. Dat is onderzoek waarbij het mogelijke effect van de maatregel wordt bepaald door het aantal slachtoffers of ongevallen vóór invoering van de maatregel te vergelijken met die na invoering, zonder de omstandigheden bij invoering van de maatregel experimenteel te manipuleren. Als over langere periodes data worden verzameld kan gebruik worden gemaakt van tijdreeks analyse, die robuuster wordt door gebruik te maken van toevoeging van versturende variabelen (indien gekend en meten), en vergelijking met een referentiegroep.

Een andere mogelijkheid is het gebruik van de Empirical Bayes methode, waarbij wordt gecorrigeerd voor het regressie naar het gemiddelde effect. Dit gebeurt door gebruik te maken van gegevens in een referentiegroep, de gegevens in de experimentele groep in de voorperiode daarvoor te corrigeren, en dit te vergelijken met de gegevens in de naperiode.

Voor de toepassing van de methodes zijn zowel verkeersonveiligheidsdata (slachtoffers en ongevallen) als data over mate van invoering van het beleid, als veranderingen op basis van ingevoerd beleid (denk aan gedragsveranderingen zoals snelheidsgedrag, gordeldracht of alcoholgebruik) nodig.

Als niet kan worden beschikt over betrouwbare ongevalldata kunnen data worden verzameld over gedragingen die met ongevallen verband houden, zoals conflicten, snelheden of bepaalde soorten bekeuringen. Afhankelijk van de evaluatievraag en de beschikbaarheid van informatie kan worden gekozen voor een kwantitatieve of kwalitatieve methode van evalueren of een combinatie van beide. Door middel van triangulatie kan de betrouwbaarheid en validiteit van data en daarop gebaseerde uitspraken worden vergroot.

Geconcludeerd wordt dat het uitvoeren van een ex post evaluatie moet worden afgewogen ten opzichte van de noodzaak en diepgang van verantwoording, de mogelijkheid om van de evaluatie te leren, praktische (on)mogelijkheden om een evaluatie uit te voeren en kosten.

Bij het uitvoeren van een ex post evaluatie in een natuurlijke setting is het vrijwel onmogelijk om waterdichte conclusies te trekken. Er zijn wel mogelijkheden om de plausibiliteit van werking van de ingevoerde maatregel te vergroten door gebruik te maken van langere tijdreeksen van observaties, te corrigeren voor versturende factoren, en verschillende soorten analyses uit te voeren.

Per onderwerp zal moeten worden overwogen of er voor een ex post evaluatie wordt geopteerd, wat de aard en de diepgang van deze evaluatie moet zijn, en welke methodologische aanpak nodig en mogelijk is.

2. Inleiding

2.1 Aanleiding voor dit onderzoek

Er zijn verschillende redenen om onderzoek te doen naar de mogelijkheden van ex post evaluatie van verkeersveiligheidsmaatregelen. De belangstelling vanuit beleid is groeiende, wat blijkt uit verschillende documenten.¹

2.2 Stand van zaken

In eerdere rapporten (Berveling et. al., 2009, Hanemaayer, 2010) is geconstateerd dat ex post beleidsevaluatie nog weinig ontwikkeld is. Als belangrijkste oorzaken worden genoemd:

- Vanuit de politiek-bestuurlijke en beleidscontext speelt de wens om vooruit te kijken een belangrijke rol. Het ligt in de aard van beleidsmakers om bezig te zijn met de toekomst, niet met het verleden.
- Bepaalde psychologische processen staan de wil tot evalueren in de weg. Bij de start van een project kan een 'optimism bias' een rol spelen. Betrokkenen hebben (ex ante) vaak een te optimistisch beeld van een project. Later verzet men zich tegen alles dat het positieve beeld verstoort. Meer in het bijzonder kan een negatieve evaluatie leiden tot reputatieschade.
- Organisatorische belemmeringen kunnen de praktische uitvoering van evaluaties in de weg staan. Projectdirecties worden vaak direct 'na het knippen van het lint' opgeheven. Bovendien kan gebrek aan geld en capaciteit er toe leiden dat ex post evaluaties worden verdrongen door meer urgente zaken. Daarnaast kan het natuurlijk personeelsverloop, in combinatie met de lange doorlooptijden van infrastructuurprojecten, leiden tot verwatering van kennis en expertise.
- De belangrijkste methodologische problemen zijn dat het moeilijk is om effecten van projecten te isoleren en om te weten hoe de wereld er zonder het project had uitgezien. Daarnaast is het moment van evalueren een punt dat aandacht verdient.

2.3 Doel van het onderzoek

Het doel van dit onderzoek is te schetsen wanneer een ex post evaluatie nodig is, aan te geven welke methoden, activiteiten en inspanningen dat vereist en aan te geven onder welke voorwaarden de methode of methoden voor het bepalen van de slachtofferbepaling toepasbaar is of zijn. Eventueel wordt ook de methode voor het bepalen van de kosteneffectiviteit behandeld. Aangetekend wordt dat andere indicatoren dan aantallen slachtoffers in dit rapport grotendeels buiten beschouwing zijn gelaten.

2.4 Onderzoeksvragen

Op basis van bovenstaand onderzoeksdoel zijn de volgende onderzoeksvragen geformuleerd:

1. Hoe kunnen we 'ex post' evaluatie definiëren?
2. Wat mogen we wel en niet van een ex post evaluatie verwachten? Wanneer is het zinvol om een ex post evaluatie uit te voeren? Wat zouden de doelen van een ex post evaluatie kunnen zijn?

¹ Het gaat hier om het Strategisch plan verkeersveiligheid, het Actieprogramma Verkeersveiligheid, de toezegging en de brief van de Minister in het Algemeen overleg. De uitwerking is te vinden in bijlage 2.

3. Hoe zou een methode die de besparing van slachtoffers en eventueel de kosteneffectiviteit achteraf kan evalueren (i.e. een ex post evaluatie) eruit kunnen zien?
4. In welke gevallen is het mogelijk om een zinvolle ex post evaluatie uit te voeren?
5. Wanneer wordt voor welk type en welke methode van ex post evaluatie gekozen?
6. Welke conclusies kunnen worden getrokken?

In de navolgende hoofdstukken zal op de onderzoeksvragen worden ingegaan. In hoofdstuk 3, begripsbepaling, wordt het kader geschetst waarbinnen ex post evaluatie kan worden geplaatst, en wordt het begrip nader uitgewerkt (vraag 1). In hoofdstuk 4 wordt ingegaan op wat van ex post evaluatie op het gebied van de verkeersveiligheid mag worden verwacht, en wat wel en niet mogelijk en doelmatig is (vragen 2 en 4). In hoofdstuk 5 wordt op kwantitatieve methodes van evalueren van verkeersveiligheidsmaatregelen ingegaan (vraag 3), en in hoofdstuk 6 wordt een beschouwing over zinvolheid van ex post evaluaties gegeven (vraag 5). In hoofdstuk 7 tenslotte, worden conclusies getrokken (vraag 6).

3. Begripsbepaling

In dit hoofdstuk worden kernbegrippen rond beleidsevaluatie toegelicht, om het kader te schetsen waarbinnen dit onderzoek is verricht.

3.1 Evaluaties in fasen van het beleidsproces

Onder ex post evaluatie wordt verstaan de vorm van evaluatie die na afloop van een project, activiteit of proces plaatsvindt.

Ex post evaluatie onderscheidt zich van ex ante evaluatie die vooraf gebeurt om de verwachte effecten te schatten, en van ex durante evaluatie die tijdens het proces van invoering wordt gepleegd. Ex durante evaluatie gebeurt vooral om te bezien of en in hoeverre de geplande activiteiten of tussenresultaten worden gerealiseerd, om daarop eventueel bij te sturen. Daarin heeft het een monitorfunctie. Bij ex durante speelt de procesevaluatie een belangrijke rol.

In het algemeen gaat het bij de evaluatie ex post om twee hoofdvragen:

- is het beoogde beleid geïmplementeerd, en in welke mate is dat het geval geweest, dan wel hoe is de implementatie voorbereid (procesevaluatie); bij verkeersveiligheid betreft dit vooral de vraag of alle betrokken partijen meedoen, en of alle afspraken zijn nagekomen;
- is het beoogde resultaat behaald, en waaraan is dat toe te schrijven?

Bij verkeersveiligheid kan het eindresultaat worden uitgedrukt in bespaarde slachtoffers door de getroffen maatregel of het maatregelenpakket.

Ook kan het resultaat worden gegeven in termen van de mate waarin is voldaan aan de te leveren inspanningen, waarvan mag worden aangenomen dat ze bijdragen aan verkeersveiligheid, bijvoorbeeld uren politie-inzet of het bereik van een campagne.

3.2 Beleidsevaluatie

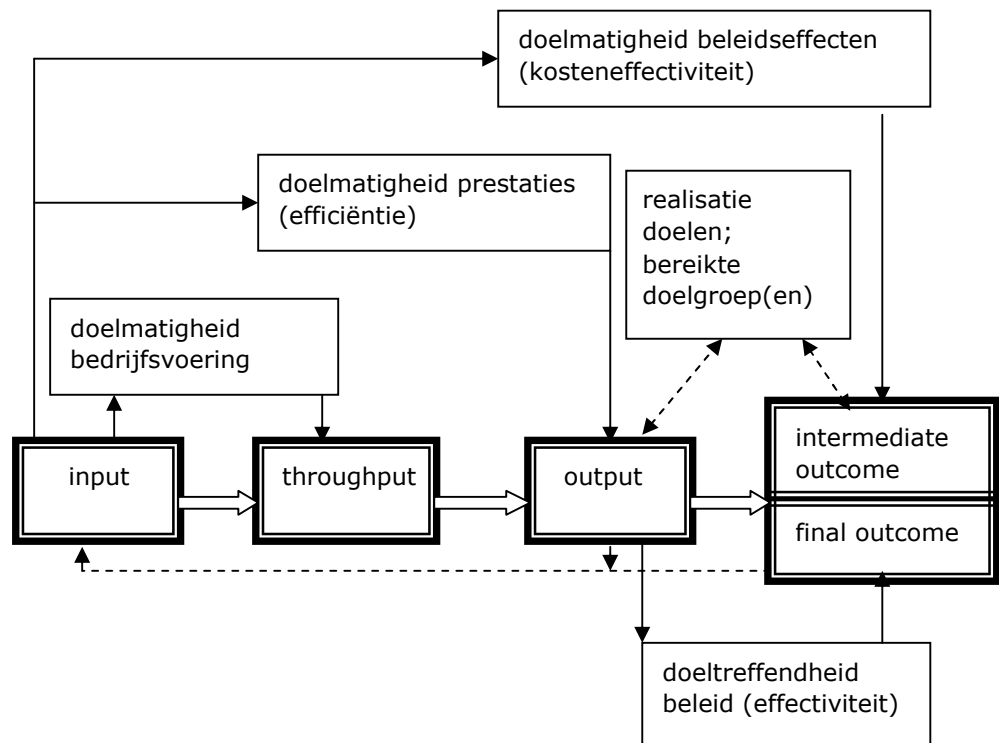
Beleidsevaluatie mag zich verheugen in een toenemende belangstelling (De Peuter et. al., 2007, Hanemaayer, 2010). Beleid moet zowel vooraf (ex ante) als na afloop (ex post) kunnen worden verantwoord. De verantwoordingsfunctie is belangrijker geworden onder invloed van een toenemende verantwoordingsvraag van de burger over activiteiten van de overheid en vragen van de Tweede Kamer over versnelling en verbetering van verantwoording van beleid (De Waal en Kerklaan, 2004).

Bij ex post beleidsevaluatie wordt onderscheid gemaakt in:

- beleidsdoorlichting, de evaluatie van beleid op het niveau van de algemene of operationele doelstellingen;
- effectenonderzoek, de evaluatie van beleidsresultaten.

Beleidsevaluatie kan betrekking hebben op verschillende aspecten van het productieproces in de publieke sector. Dit wordt weergegeven in het schema op de volgende bladzijde (Algemene Rekenkamer, 2005; Wouters et. al., 2009). Dit schema is ook van toepassing op de evaluatie van verkeersveiligheidsbeleid.

Schema 3.1 Relatie evaluatie-aspecten en beleidsproces



Toelichting bij schema 3.1:

- input betreft de inzet van mensen en middelen. In de verkeersveiligheidssector wordt doorgaans onderscheid gemaakt in de 3 E's + O: engineering (maatregelen aan fysieke omgeving of voertuig), education (scholing en voorlichting), enforcement (wetgeving en handhaving) en organisatie van beleid.
- throughput betreft processen en procedures bij het inzetten van de input; de kwaliteit van de throughput wordt gemeten door middel van een procesevaluatie.
- output zijn de producten die worden opgeleverd zoals een rotonde, realisatie dode hoek spiegel, uren inzet politie, een verkeersmethode, een organisatiemodel.
- outcome betreft de beoogde resultaten, in termen van gedrag (intermediate outcome) en ongevalsreductie (final outcome).

Voorbeeld van indicatoren voor de effectiviteit van verkeersveiligheidsbeleid:

Indicatoren voor intermediate outcome

- rijsnelheden
- gebruik alcohol, drugs, medicijnen
- gebruik beschermingsmiddelen
- rood licht negatie

Indicatoren voor final outcome

- verkeersdoden
- ernstig gewonden
- licht gewonden
- ongevallen
- verloren levensjaren
- maatschappelijke kosten verkeers-
onveiligheid

De output of outcome kan basis zijn voor een nieuwe input, bijvoorbeeld als op basis van gegevens een beleidsmaatregel wordt bijgesteld of uitgebreid. De evaluatiecyclus wordt in de loop der tijd meerdere malen doorlopen.

De volgende aspecten en relaties zijn in het schema op blz. 12 opgenomen:

1. doelrealisatie. Hierbij worden de doelen van beleid gelegd naast de situatie nadat beleid is doorgevoerd; zijn de beoogde doelen inderdaad behaald.
2. doelgroepbereik. De vraag of de personen die beïnvloed moeten worden daadwerkelijk met beleid in contact zijn gekomen.
3. doelmatigheid of efficiëntie van beleid. Hierbij gaat het om de vraag tegen welke kosten of inspanningen beleidsresultaten zijn geleverd, en of dat beter kan. De volgende aandachtspunten zijn van belang:
 - a. hadden dezelfde effecten of prestaties met minder middelen gerealiseerd kunnen worden (inputoriëntatie)
 - b. was met dezelfde inzet van middelen meer mogelijk geweest (output-/outcomeoriëntatie).
4. doeltreffendheid van beleid. Hierbij gaat het om de vraag in hoeverre de gewenste beleidsresultaten, prestaties of effecten zijn toe te schrijven aan het gevoerde beleid. Was dat effectief?

Bij de evaluatie gaat het om de relaties tussen:

- ingezette middelen (personeel, materieel, geld) en proces (input en throughput); is er *economisch* met de middelen omgegaan?
- ingezette middelen (input), proces (throughput) en output (producten, diensten, activiteiten); is de *bedrijfsvoering doelmatig* geweest?
- ingezette middelen, proces, output en uitkomst (van input tot en met outcome), ofwel de effecten zoals bedoeld; is het *beleid doelmatig* (kosteneffectief) geweest?
- output en outcome; dit betreft de vraag naar *doeltreffendheid* van beleid.

In het proces van beleidsuitvoering zijn er verschillende meetmomenten om de doelmatigheid en de doeltreffendheid van beleid vast te stellen. De meetmomenten worden mede bepaald door het tijdstip waarop beleidsbesluiten worden genomen.

3.3 Benaderingswijzen

Bij ex post evaluatie kan gebruik worden gemaakt van kwantitatieve en kwalitatieve aanpakken (Wouters et.al, 2009). In het algemeen wordt van kwalitatief onderzoek gesproken als de aantallen onderzoekseenheden (proefpersonen, locaties, te onderzoeken voertuigen) beperkt zijn en als de nadruk ligt op exploratie van het onderwerp van onderzoek (Van der Zee, 2004). Kwalitatief en kwantitatief onderzoek kunnen elkaar uitstekend aanvullen. Op de dimensie kwantitatief naar kwalitatief kunnen de volgende benaderingswijzen worden onderscheiden (Hanemaayer, 2010):

- evidence based benadering, waarbij de nadruk ligt op het meten van de veranderingen die aan beleid kunnen worden toegeschreven. Verklaringen van effecten kunnen daardoor minder nadruk krijgen, en ook is er het risico dat te weinig recht wordt gedaan aan moeilijk meetbare factoren;
- realistic evaluation, waarbij het inzicht in verklaringen juist benadrukt wordt; wat gebeurt er in de werkelijkheid?;
- complexity evaluation, waarbij in de complexe werkelijkheid rekening wordt gehouden met externe omstandigheden. Omdat de causaliteit moeilijk is vast te stellen komt het leren van evaluatie voorop te staan;

- onderhandelend evalueren, waarbij de evaluator de rol krijgt van facilitator, en het evalueren een proces van samenspraak wordt. Daardoor fungeren betrokken partijen als partners die meteen van de evaluatie kunnen leren.

Bij kwantitatieve methoden kunnen de volgende typen worden onderscheiden (Wouters et al, 2009):

- meten van intermediate of final outcome via longitudinaal onderzoek of puntmetingen (vergelijkend onderzoek).
- schatten op basis van expertmeningen.
- modelleren met quasi experimentele onderzoeksmethode.

Voor de nadere uitwerking wordt verwezen naar hoofdstuk 5.

Bij kwalitatief onderzoek wordt nagegaan hoe de relaties tussen wezenlijke variabelen elkaar beïnvloeden. Belangrijke methoden zijn:

- exploratief of beschrijvend onderzoek dat is gericht op het verkennen van ideeën over mogelijke samenhangen, of het genereren van hypothesen. Een case study is het meest basaal, andere exploratieve onderzoeken betreffen een beperkt aantal onderzoekseenheden.
- meta evaluatie, waarbij een oordeel wordt gegeven over de kwaliteit van een evaluatie of reeks evaluaties. De criteria zijn meestal betrouwbaarheid, geloofwaardigheid, en bruikbaarheid van bevindingen.
- peer review, of intercollegiale toetsing, waarbij liefst onafhankelijke experts hun visie geven op gepresenteerde cases of gegevens.

Veel gebruikte werkwijzen bij kwalitatief onderzoek zijn: vragenlijst-onderzoek, interviews, participierend observeren, discussiegroepen, pannels, analyse van verhalen, reconstructie van gebeurtenissen, beschrijven van ervaringen, gebruik van het detective paradigma (zie kader). Zulke werkwijzen kunnen ook in kwantitatief onderzoek worden gebruikt.

Het detective paradigma

Dit is een bestuurskundige evaluatiemethode waarbij de evaluator de effectiviteitsvraag van maatregelen aanpakt, als ware hij een detective die een moordzaak moet oplossen. Er worden hypothesen gesteld over de wijze waarop de inzet van een maatregel of pakket maatregelen kan hebben geleid tot reductie van het aantal slachtoffers. De relaties van maatregel via intermediate outcome naar final outcome worden stuk voor stuk getoetst (Van 't Hoff et al, 2000).

Om de robuustheid van resultaten te vergroten kan gebruik worden gemaakt van triangulatie, het vergelijken van gegevens met tenminste twee andere bronnen. Er wordt onderscheid gemaakt in data triangulatie, waarbij data uit verschillende bronnen worden vergeleken, methodische triangulatie waarbij van verschillende dataverzamelings- en verwerkingstechnieken gebruik wordt gemaakt, onderzoekstriangulatie waarbij verschillende onderzoekers verschillende technieken toepassen, en theoretische triangulatie waarbij de resultaten met verschillende eerdere bevindingen en gezichtspunten worden vergeleken.

3.4 Ex ante beleidsevaluatie

Beleidsverantwoording achteraf vergt zorgvuldige afwegingen vooraf. Hierin kan een goede ex ante evaluatie voorzien. Volgens een publicatie van het ministerie van Financiën (2003) heeft evaluatie ex ante drie hoofddoelen:

1. inzicht krijgen in de effecten van beleidsvoornemens, te leveren prestaties en in te zetten middelen;
2. richting geven aan strategische keuzes bij beleidsvoorbereiding;
3. achteraf systematisch verantwoording over gevoerd beleid afleggen.

Het derde doel slaat de brug met ex post evaluatie. Ex post kan worden nagegaan of beleid is uitgevoerd zoals was beoogd, en of de voorspelde effectiviteit en efficiëntie van beleid is waargemaakt.

Daarom is het van belang om bij de ex ante evaluatie goed te overdenken welke indicatoren ten aanzien van doelrealisatie worden gekozen, de vragen voor ex post evaluatie in gedachten houdend.

Bij ex ante evaluatie kan onderscheid worden gemaakt in integrale en partiële ex ante onderzoeken. Bij integrale onderzoeken worden alle verwachte effecten en neveneffecten in kaart gebracht. Bij partiële onderzoeken kunnen bepaalde onderwerpen aan de orde worden gesteld, bijvoorbeeld effecten op het milieu (MER), op emancipatie, veiligheid of bedrijfsmatige opbrengsten.

Ook kan onderscheid worden gemaakt in evaluatie van afzonderlijke maatregelen en van een totaal beleidspakket.

Methoden voor ex ante evaluatie zijn:

- Multi-criteria analyse; hierbij wordt een serie beleidsalternatieven gewogen op verwachte effecten, die niet alle in geld zijn uit te drukken;
- Kosteneffectiviteitsanalyse (KEA); hierbij wordt onderzocht met welk beleidsalternatief een bepaalde doelstelling zo goedkoop mogelijk kan worden gerealiseerd, of wat de maximale realisatie is, gegeven beschikbare middelen, met in acht neming van vastgestelde kwaliteitscriteria;
- Kosten-batenanalyse (KBA); voor- en nadelen van een beleidsvoornemen worden zoveel mogelijk in geld uitgedrukt, en zo tegen elkaar afgewogen;
- Doorrekening aan de hand van een model. Verkeersmodellen worden zinvol geacht voor ex ante evaluaties waarbij scenario's van maatregelenpakketten of alternatieve systeemontwerpen met elkaar worden vergeleken op basis van verkeersveiligheidseffecten, eventueel in combinatie met andere effecten (Morsink en Wismans, 2008).

- een voorbeeld van multi-criteria analyse is te vinden in het ADVISORS project. Hier is een weging gemaakt van verschillende in car systemen op basis van een aantal criteria t.a.v. producenten, gebruikers en de maatschappij (Ampe et. al., 2008).
- een voorbeeld van kosteneffectiviteitsanalyse is het project "bouwstenen voor een veiligheidsambitie". Hierin zijn de verkeersveiligheidseffecten en de kosten van verschillende infrastructurele maatregelscenario's voor het rijkswegennet naast elkaar gezet. Per type maatregel is de kosteneffectiviteit bepaald: wat kost de maatregel en wat levert het op in termen van besparing van slachtoffers (Beenker en Verdiesen, 2010).
- voorbeeld van een kosten-batenanalyse is de berekening van kosten van invoering van schadepreventieprogramma's bij transportbedrijven en van boordcomputers in vrachtwagens en de slachtofferbesparing die dit oplevert, uitgedrukt in geld (Langeveld & Schoon, 2004).

Ex ante evaluatie van beleid is niet altijd vereist, maar moet wel altijd worden overgewogen. Bij verkeersveiligheidsbeleid is er vrijwel altijd sprake van een ex ante evaluatie.

Op basis van ex ante evaluatie wordt beleid vastgesteld. Daarbij kan besloten worden om een maatregel of maatregelpakket te implementeren. Ook kan worden besloten tot pilots, met name als de ex ante inschattingen grote onzekerheid hebben, of als er verschillende scenario's worden voorgesteld. Op basis van ex ante evaluatie kan worden voorzien in volgende evaluatiemomenten, middelen en aanpakken.

Een voorbeeld van een pilot is de maatregel bromfiets op de rijbaan, die in 1991 in drie gemeenten als proef is uitgevoerd. De evaluatie wees uit dat het aantal letselslachtoffers met 70% daalde, terwijl de daling in de vergelijkingsgroep 20% bedroeg (Hagenzieker, 1995). De bevindingen zijn mede aanleiding geweest om de maatregel bromfiets op de rijbaan landelijk in te voeren in december 1999. Bij de maatregel bromfiets op de rijbaan is voorzien in een ex post evaluatie, die is uitgevoerd in 2000 (Van Loon, 2001).

Het is belangrijk om de ex ante aanpak goed te documenteren, zodat tijdens implementatie door middel van procesevaluatie kan worden bijgesteld en zodat bij ex post evaluatie zicht is op de gestelde doelen, de aanpak, de relevante output, meetcriteria en besluiten tijdens de uitvoering van een project of maatregel die hebben geleid tot aanpassingen van de oorspronkelijke planning of uitvoering.

3.5 Ex durante beleidsevaluatie

Ex durante evaluatie vindt vaak plaats als bepaalde mijlpalen bij de implementatie van beleid zijn bereikt. De ex durante evaluaties hebben een monitoringsfunctie. Een nulmeting kan deel uitmaken van de ex durante beleidsevaluatie. Vaak wordt gepoogd een actueel inzicht te verkrijgen in het doelbereik, om van ervaringen te leren, en aanbevelingen te doen voor eventuele bijsturing van beleid, nog tijdens de uitvoering van de maatregel of het project. Bij ex durante wordt veelal gebruik gemaakt van procesevaluatie; is de opzet en uitvoering verlopen zoals gepland, en zo nee, wat is niet gerealiseerd of veranderd, en waarom? Bij de opzet van ex durante evaluaties is het van belang om de indicatoren en meetinstrumenten goed te documenteren, omdat deze in vervolgevaluaties en de ex post evaluatie als ijkpunt kunnen fungeren. Ex durante evaluatie kan tot uitbreiding, verfijning of verbetering van het meetinstrumentarium leiden, als blijkt dat betrouwbaarheid of validiteit tekort schieten. Dit heeft gevolgen voor de werkwijze bij de ex post evaluatie.

3.6 Ex post beleidsevaluatie

Met ex post beleidsevaluatie moet de klassieke vraag worden beantwoord of de overheid de goede dingen doet, en of ze de dingen goed doet. Doet het beleid wat is afgesproken, en wat zijn de resultaten? Evaluatie ex post kan worden gedefinieerd als systematisch onderzoek naar beleid, de wijze waarop beleid is uitgevoerd, en de effectiviteit, kosten en kwaliteit van geleverde producten en diensten. Het gaat om het vaststellen van de doelmatigheid en doeltreffendheid van beleid. Zoals Van der Vlist et. al. (2007) aangeven worden met ex post evaluaties in essentie drie vragen beantwoord:

- wordt het gestelde beleidsdoel of de gestelde beleidsdoelen gehaald?
- heeft het beleid hier inderdaad een bijdrage aan geleverd?
- was de bijdrage van beleid effectief?

De vragen kunnen worden beantwoord door na te gaan of de ex ante veronderstellingen juist zijn gebleken.

Ex post beleidsevaluatie vindt in de beleidscyclus plaats na de beleidsvoorbereiding, beleidsbepaling en beleidsuitvoering, en voor de terugkoppeling.

Door Berveling et. al (2009), wordt opgemerkt dat het van belang is, gezien de complexiteiten, om tijdig bij de vaststelling van beleid ook een kader voor ex post evaluatie vast te stellen. Daarbij laten zich de volgende vragen stellen:

- a. wordt ex post evaluatie ingezet om te leren of om te verantwoorden? Beide aspecten zijn doorgaans van belang;
- b. moeten alle projecten en alle aspecten worden onderzocht, of volstaat een selectie;
- c. op welke termijn na beleidsinvoering geeft een meting het meest optimale resultaat;
- d. op welke niveaus moet worden geëvalueerd, en kan worden volstaan met een lichte evaluatie of is een middelzware of zware evaluatie vereist?

De antwoorden op deze vragen hangen samen met de door het beleid beoogde opbrengst van ex post evaluatie. Bij ex post evaluatie van overheidsbeleid kan het om vier hoofdvragen gaan.

1. is het doel of zijn de doelen bereikt? Zijn daarbij de beoogde doelgroepen bereikt en zijn de beoogde maatregelen uitgevoerd conform de plannen? Als dat niet het geval is, wat is er dan gewijzigd, en waarom?
2. is het al dan niet bereiken van het doel of de doelen toe te schrijven aan het gevoerde beleid en de gebruikte instrumenten? Bij deze vraag speelt de causaliteitsproblematiek een belangrijke rol: hoe weet je of en in welke mate de geconstateerde resultaten daadwerkelijk door de beleidsinterventie tot stand zijn gekomen, en niet als gevolg van andere ontwikkelingen? Dit probleem kan kwantitatief worden aangepakt, door via een zuiver experiment met een aselekt gekozen experimentele en controlegroep te werken (zie hoofdstuk 5). Ook kan het kwalitatief worden aangepakt, door aan de hand van eerdere ervaringen, theorieën of eliminatiemethode plausibel te maken dat de belangrijkste werkende factoren inderdaad de bedoelde maatregelen zijn.
3. is het beleid doelmatig? Wegen de effecten op tegen de kosten? Hiervoor wordt een kosteneffectiviteitanalyse uitgevoerd, waarbij de gerealiseerde effecten worden vergeleken met de benodigde investeringskosten. Bijvoorbeeld welke veiligheidswinst staat tegenover de kosten van invoering van het praktijkexamen voor bromfietzers?
4. is de beleidsuitvoering doelmatig; ofwel hoe wordt de verhouding van kosten en kwaliteit van prestaties, producten en diensten beoordeeld? Hadden met dezelfde middelen meer effecten bereikt kunnen worden, of hetzelfde effect met minder middelen? Hiervoor wordt doorgaans een kostenbaten analyse uitgevoerd, waarbij de effecten zoveel mogelijk in geld worden uitgedrukt. Bijvoorbeeld was een praktijkexamen met drie bromfietzers tegelijk even effectief geweest als met 1 bromfietser, en hoeveel goedkoper was dat geweest?

3.7 Beschouwing

Bij verkeersveiligheidsbeleid wordt ex post getoetst wat de bijdrage is geweest van het beleid in de verkeersveiligheidsontwikkelingen. Afhankelijk van het doel en de context van ex post evaluatie kan op verschillende niveaus worden geëvalueerd. Soms volstaat evaluatie op intermediate outcome niveau, of evaluatie op het doeltreffendheids criterium.

Bijvoorbeeld bij de meting van effecten van autogordels, rijden onder invloed of snelheid wordt de relatie met ongevallen bekend verondersteld, en dan kan worden volstaan met metingen van het gedrag zelf.

Verkeersveiligheidsbeleid speelt zich af in de complexe werkelijkheid, waarin sprake is van een groot aantal elkaar onderling beïnvloedende factoren. In het algemeen wordt een pakket aan maatregelen parallel in de tijd ingevoerd, bijvoorbeeld nieuwe regelgeving, uitvoering, technieken, opleiding en handhaving. Het is lastig, zo niet onmogelijk om precies de bijdrage van elke factor te onderscheiden, zeker wanneer er interactie is tussen verschillende maatregelen.

Ook speelt het onderscheid tussen maatregelen die op één welbepaald moment worden ingevoerd (zoals de helmplicht), en maatregelen die over langere tijd worden gerealiseerd, zoals het ombouwen van conventionele kruispunten naar rotondes. Bij dit laatste is het nog moeilijker om precies die factoren te isoleren waarop het beleid is gericht, omdat er tijdens implementatie andere zaken in de realiteit veranderen.

In hoofdstuk 5 wordt ingegaan op diverse onderzoeksmethodes, hun voorwaarden en hun grenzen.

4. Doelen en betekenis ex post beleidsevaluatie

In het vorige hoofdstuk is uitgebreid ingegaan op verschillende aspecten van beleidsevaluatie. Daarbij is een onderscheid gemaakt in de doelen "leren" en "verantwoorden". In dit hoofdstuk wordt eerst het doel verantwoorden nader beschouwd (par. 4.1), en vervolgens het leren (par. 4.2). Een andere manier om naar het doel van evaluatie te kijken is verbonden met de fase van evalueren. Dit wordt uitgewerkt in par. 4.3. Tenslotte wordt een beschouwing gegeven over de betekenis van ex post evaluatie in par. 4.4.

4.1 Verantwoorden

Verantwoorden kent verschillende aspecten:

- inhoudelijke verantwoording, waarbij wordt beschreven welke algemene aanpak en middelen voor de afgesproken beleidsdoelen zijn ingezet, en wordt aangegeven waarom voor deze aanpak en middelen is gekozen. Op grond daarvan kan worden bepaald of de effectiviteit voldoende was (in hoeverre zijn de doelen behaald), en of inderdaad de inzet aan de afgesproken doelen ten goede is gekomen.
- procedurele verantwoording. Hierbij wordt nagegaan of conform een afsprakensysteem (juridisch, contractueel of volgens de regels van een bepaalde discipline) is gehandeld. De regels moeten expliciet en nauwkeurig worden beschreven. Op grond daarvan kan worden nagegaan of er bij de uitvoering procedureel correct en integer is gehandeld. Zeker wanneer men aansprakelijk gesteld zou kunnen worden is procedurele verantwoording van belang, maar algemeen geldt dat beleid volgens bepaalde afspraken moet worden uitgevoerd.
- financiële verantwoording; dit wordt verlangd voor financiering of subsidiering of voor de accountantsverklaring bij jaarrekeningen. Financiële verantwoording vergt dat alle input, activiteiten en output zoveel mogelijk in geldwaardes wordt uitgedrukt, zoals bij de KBA. Vervolgens moet er een beoordelingskader zijn om na te gaan of inkomsten op verantwoorde wijze zijn gegenereerd en uitgaven verantwoord zijn besteed. Op grond daarvan kunnen uitspraken worden gedaan over de vraag of de middelen op de juiste wijze zijn ingezet.

Het verantwoorden is nodig om besluiten te nemen over de beleidssturing: is het wenselijk om het gevoerde beleid voort te zetten, te versterken, af te bouwen of zelfs stop te zetten (indien beleidsmatig mogelijk)?

Bij de uitvoering van de evaluatie moet de gekozen evaluatie-aanpak eveneens worden verantwoord. Beschreven wordt welke definities, indicatoren, methoden en technieken, meetinstrumenten, analysetechnieken en beschrijvingsvormen zijn gehanteerd bij de evaluatie. Daarbij wordt zo gedegen mogelijk de betrouwbaarheid en validiteit van instrumenten beschreven. Hierbij speelt de overweging dat er een proportionaliteit moet zijn tussen de gekozen evaluatie aanpak en de te verwachten uitkomsten of besluiten die op basis van de uitkomsten genomen worden.

Verantwoorden is belangrijker als er sprake is van:

- een controversiële maatregel of maatregelenpakket,
- een maatregel met een grote maatschappelijke impact,
- twijfel aan het effect van de maatregelen (op basis van ex ante schattingen),
- hoge kosten van het maatregelenpakket,
- uitbreiding of aanpassing van het maatregelenpakket in de toekomst,
- groot belang van toetsing van de haalbaarheid.

Bij het verantwoorden is van belang dat vooraf duidelijkheid bestaat over de belangrijkste beleidsvragen die beantwoord moeten worden. In de ex ante evaluatie moeten deze vragen zoveel mogelijk zijn onderzocht.

Van een andere orde is de vraag of verantwoorden altijd mogelijk en nodig is, en daarmee de vraag of verantwoorden altijd doel van de evaluatie moet zijn. Als een maatregel zeker ingevoerd wordt (zoals de aanleg van een nieuwe verbindingsweg waarover al is besloten) dan heeft verantwoorden van de maatregel als zodanig weinig zin, omdat de maatregel zelf toch niet meer wordt teruggedraaid. Over vormen van uitvoering kan overigens wel verantwoording worden afgelegd, om na te gaan of het meest veilige toepassingsalternatief is gekozen. Als het gaat om verwachte kleine effecten in een complexe setting dan is de vraag of verantwoorden mogelijk is, vanwege methodologische problemen.

4.2 Leren of kennisgeneratie

Ook het leren kent verschillende aspecten:

- heroverwegen van de doelen, het proces of de producten. Evaluatie kan opleveren dat er beleidsmatig onvoldoende duidelijkheid heeft bestaan over doelen, proces of verwachte resultaten. Volgens Van der Vlist et. al. (2007) komt het regelmatig voor dat er te weinig inzicht is gegeven in de beleidsproblematiek, mede vanwege de complexiteit van deze problematiek. Heroverweging van beleid kan een uitkomst zijn van de ex post evaluatie. Dit heroverwegen kan leiden tot aanscherping of bijsturing waardoor vervolgactiviteiten een andere richting of investering krijgen.
- kennisopbouw voor analoge of verwante beleidsprocessen. Door ex post te evalueren kunnen verwante beleidsprocessen beter worden aangestuurd, of als ze al in gang zijn gezet, nog worden bijgestuurd. Dit vereist dat kennis van ex post beleidsevaluaties toegankelijk en up to date is.
- kennisopbouw voor toekomstige aanpakken. Ervaringen worden gedocumenteerd om soortgelijke projecten of activiteiten in de toekomst te vergemakkelijken of te behoeden voor missers ("leren van fouten"). Het komt nu regelmatig voor dat men het wiel opnieuw uitvindt, en vermijdbare fouten maakt omdat er te weinig documentatie of kennisuitwisseling is over al uitgevoerde vergelijkbare beleidsactiviteiten en de resultaten ervan.
- leren ten behoeve van het evalueren zelf. Het uitvoeren van ex post beleidsevaluaties kan er toe leiden dat kennis over de aanpak van dit type evaluatie wordt uitgebreid of verfijnd. Daardoor kan de uitvoering van de evaluatie beter verlopen en tot robuustere resultaten leiden. Ervaringen met het uitvoeren van evaluaties hebben geleid tot methodes van onderzoek en analysetechnieken waardoor meer grip wordt verkregen op het bepalen van de effecten van de geplande interventie (zie ook hoofdstuk 5).

Nederland behoort tot de top van verkeersveilige landen. Binnen Nederland is sprake geweest van kennisontwikkeling, zoals demonstratiegebieden duurzaam veilig, fietsvoorzieningen, rotondes, etc. De kennis van onder andere ons Duurzaam Veilig programma is naar vele landen geëxporteerd, waardoor de verkeersveiligheid ook in andere landen een sprong heeft gemaakt.

4.3 Proces, prestatie en product

Bij ex post evaluatie kan het doel zijn (zie ook schema 3.1):

- vaststellen hoe de uitvoering van een programma, project of maatregel verloopt of is verlopen (procesevaluatie),
- vaststellen van output (prestatie-evaluatie),
- vaststellen van intermediate of final outcome (productevaluatie).

Het accent kan op elk van de drie doelen liggen. Er moeten doorgaans afwegingen worden gemaakt over waar het accent komt te liggen. Dit is afhankelijk van de context van evaluatie, het onderwerp, de aard van besluitvorming, reeds voorhanden kennis en data, en beschikbare middelen om te evalueren.

Vaak geldt dat hoe ingrijpender, gevoeliger, onbekender en zichtbaarder een maatregel is, hoe zwaarder de evaluatie wordt aangezet en hoe meer aspecten er worden gemeten.

De invoering van 30 km/u gebieden kon op een groot draagvlak rekenen, zowel bij beleid als bij de burgers. In pilot studies (Eindhoven en Rijswijk) is gebleken dat er sprake is van draagvlak en een hoge effectiviteit. Dat laatste kan ook worden afgeleid uit de relatie snelheid – ongevallen - afloop van ongevallen op basis van onderzoek van onder andere Nilsson. In het geval 30 km/u kan dan bij evaluatie het accent komen liggen op proces en output; zijn processen verlopen zoals beleid voor ogen had (in dit geval met name de uitvoering door decentrale overheden), en zijn de gebieden ingericht zoals beoogd. Op deze basis kan beleid afdoende volgende besluiten nemen.

Bij een nieuwe ingrijpende maatregel, zoals de spitsstroken en plusstroken wordt ook naar de final outcome gekeken. Bij deze maatregel speelt uitdrukkelijk de conditie dat de veiligheid niet mag verslechteren. Dit vraagt een degelijke ex post evaluatie op het niveau van slachtoffers, en daarvoor moet een goed design worden ontwikkeld.

Bij deze maatregel speelt zowel het aspect van leren als van verantwoorden een rol, omdat over een aantal uitvoeringsvarianten nog vragen zijn, en omdat beleid de besluiten over spits- en plusstroken moet waarmaken.

Een duidelijke beleidscontext maakt zinvolle keuzes mogelijk in accenten bij ex post evaluatie. Met name inzicht in de speelruimte bij besluitvorming en de aard van mogelijke besluiten is voorwaardelijk bij het overwegen van de vorm van ex post evaluatie. Er moet doelmatig en verstandig met de verschillende mogelijkheden van ex post evaluatie worden omgegaan. Dat betekent dat de benodigde middelen voor evaluatie moeten passen bij het doel van de evaluatie en de beschikbare informatie. Besloten moet worden wanneer een "zware" en wanneer een "lichtere" vorm van evalueren van toepassing is.

4.4 Betekenis ex post beleidsevaluatie

In de beleidscyclus vinden evaluaties op verschillende momenten plaats.

- in de beleidsvoorbereidende en de beleidsbepalende fase, waarin de situatie wordt geanalyseerd en planning en uitvoering worden vastgesteld, worden ex ante evaluaties gehanteerd om het beleid zo goed mogelijk te funderen;

- in de beleidsuitvoerende fase, waarin projecten of maatregelen worden uitgevoerd, is bijsturing van groot belang, en daar geeft procesevaluatie input voor;
- in de beleidsevaluatiefase leidt ex post evaluatie tot een conclusie over het gevoerde beleid en tevens kunnen aanwijzingen worden gegeven voor verbetering van toekomstig beleid.

Om de functies van ex post evaluatie zo goed mogelijk te waarborgen moeten er voldoende middelen zijn om de evaluatie gedegen uit te voeren. Dat betekent overigens niet dat elke maatregel of elk project uitputtend, op outcome niveau ex post geëvalueerd moet worden. Er kunnen overwegingen zijn om niet te evalueren, een lichte vorm van evaluatie te gebruiken of om niet verder te gaan dan evaluatie van proces of output. Het betreft overwegingen zoals:

- hebben analoge maatregelen of projecten al voldoende informatie opgeleverd?
- ligt bij de evaluatie de nadruk op verantwoord en leren?
- staan de verwachte inspanningen of kosten van evaluatie in verhouding tot de verwachte opbrengsten?
- welke rol spelen de evaluatieresultaten bij verdere beleidsontwikkeling of besluitvorming? Wordt hierbij vooral gestuurd op processen, de output of de outcome?
- welke data en meetmiddelen zijn voorhanden?
- op welke termijn moeten resultaten worden opgeleverd?

Dergelijke overwegingen geven sturing aan aard en diepgang van de ex post evaluatie. Zij kunnen per project of maatregel verschillen, en daarom zal de afweging voor uitvoering van ex post evaluatie steeds opnieuw moeten worden gemaakt.

Ook is het belangrijk dat de resultaten van de evaluatie zodanig worden gepubliceerd en gecommuniceerd dat de analyse en conclusies transparant zijn voor de betreffende beleidsfunctionarissen. Tevens is het van belang dat zo concreet mogelijk handvatten worden geboden om het beleid aan te passen, of beter voorbereid te zijn op toekomstige beleidsinterventies. Niet altijd zijn evaluaties zo uitgevoerd of beschreven dat beleidsfunctionarissen ze als leer- en sturingsbron kunnen gebruiken. Met andere woorden, het onderzoek levert niet altijd bruikbare conclusies of aanbevelingen op. De oorzaak daarvan kan liggen in de complexiteit of het te weinig toegespitst zijn van de beleidsvragen, waardoor van een evaluatie bezwaarlijk zeer concrete resultaten verwacht kunnen worden. Een ander probleem kan zijn dat de vertaalslag van evaluatieresultaten naar de gestelde beleidsvraag lastig is, omdat er geen eenduidige resultaten worden verkregen (de werkelijkheid is doorgaans genuanceerd) of omdat het niet mogelijk is om meetresultaten een op een naar beleidshandelingen te vertalen.

Deze problemen kunnen (deels) worden ondervangen door bij het maken van de vertaalslag van onderzoek naar beleidsconclusies en beleidsaanbevelingen verschillende partijen te betrekken. Dat betreft in elk geval vertegenwoordigers van de partij die het onderzoek uitvoert, en vertegenwoordigers vanuit het beleid.

Samen zullen ze de aard en reikwijdte van gevonden resultaten moeten afwegen, om te bezien hoe deze vertaald kunnen worden naar beleidsaanbevelingen die relevant zijn voor verdere beleidsbepaling.

5. Methodologie ex post evaluatie

5.1 Inleiding

In hoofdstuk 3 is aangegeven dat ex post evaluatie kwalitatief en kwantitatief kan gebeuren.

In dit hoofdstuk wordt nader op kwantitatieve methodes van aanpak ingegaan. Een deel van de in dit hoofdstuk en bijlage 1 beschreven aanpakken is ook te vinden in CROW publicatie 248 (CROW, 2008). In het kort worden wetenschappelijke mogelijkheden besproken om verkeersveiligheidsmaatregelen ex post te evalueren. In Bijlage 1 staat een gedetailleerde verhandeling over de wetenschappelijke uitvoering van ex post onderzoek voor verkeersveiligheidsmaatregelen en de statistische onderbouwing. Deze verhandeling sluit aan op de in dit hoofdstuk behandelde methodes. Er wordt steeds verwezen naar de bijbehorende paragrafen van bijlage 1.

In dit hoofdstuk bekijken we allereerst hoe de ex post effect evaluatie van een verkeersveiligheidsmaatregel er in het ideale geval zou moeten uitzien, zodat we na afloop daadwerkelijk kunnen concluderen dat de maatregel een bepaald effect heeft *veroorzaakt*. Kort gezegd komt het erop neer dat een ideale effect evaluatie alleen kan worden uitgevoerd als we in staat zijn om de omstandigheden tijdens de invoering van een maatregel te *manipuleren*. Het is bij de evaluatie van een verkeersveiligheidsmaatregel echter meestal onmogelijk om op een experimentele manier invloed uit te oefenen op de omstandigheden waaronder de maatregel wordt ingevoerd. Naast de zuiver experimentele evaluatie bespreken we ook twee andere methodes gebaseerd op observationeel onderzoek, namelijk tijdreeksanalyse en 'empirical bayes'.

5.2 De ideale opzet – experimenteel onderzoek

De reden dat we in dit hoofdstuk eerst ingaan op de 'ideale' onderzoeksopzet is dat de principes die eraan ten grondslag liggen ons veel kunnen leren over de beste manier om het effect van een maatregel te evalueren na invoering in een natuurlijke observationele setting (zie verder bijlage 1, par. 1 t/m 3).

De ideale onderzoeksopzet voor de ex post evaluatie van de effecten van maatregelen vereist de volgende experimentele manipulaties:

1. een random trekking van de steekproef van studieobjecten uit de doelpopulatie (de populatie waarop de maatregel geacht wordt effect te sorteren);
2. een random toewijzing van de studieobjecten in de steekproef aan twee experimentele condities: een groep studieobjecten die de maatregel wel ondergaat, en een groep studieobjecten die de maatregel niet ondergaat.

De eerste manipulatie maakt dat de conclusies uit het onderzoek (welke deze ook zijn) gegeneraliseerd mogen worden naar de totale doelpopulatie omdat de random trekking garandeert dat de steekproef representatief is voor de doelpopulatie. De tweede manipulatie veroorlooft ons om te concluderen dat de maatregel een gevonden effect heeft *veroorzaakt* omdat de random toewijzing aan de twee condities garandeert dat mogelijke versturende variabelen gelijkelijk over de twee condities zijn verdeeld. Verder moet een voormeting worden gedaan, voordat de maatregel in werking gaat, en een nameting na inwerkingtreding; de verschillen tussen voormeting en nameting geven – afgezet tegen de verschillen tussen voormeting en nameting bij de groep studieobjecten die de maatregel niet ondergaat- de richting en grootte van het effect weer.

De werkelijkheid gebiedt ons te zeggen dat de ideale experimentele opzet in de praktijk zelden of nooit gerealiseerd wordt. Er zijn alternatieve experimentele designs. Die hebben gevolgen voor de conclusies die kunnen worden getrokken. Besproken worden het experiment zonder voormeting, en het experiment zonder controlegroep.

Experiment zonder voormeting

Hierbij wordt alleen gemeten na de inwerkingtreding van de maatregel. Ook nu kunnen verschillen tussen experimentele groep, waar de maatregel is ingevoerd en controlegroep zonder maatregel worden vastgesteld. Alleen als de studie objecten in de steekproef random zijn toegewezen aan de experimentele en controle conditie, mag een eventueel verschil worden verklaard als zijnde veroorzaakt door de genomen maatregel.

Experiment zonder controlegroep

Er wordt alleen bij de populatie of steekproef gekeken wat de stand van zaken is voordat de maatregel is ingevoerd, en daarna. Nu kan niet zonder meer worden vastgesteld dat het gevonden effect is toe te schrijven aan de genomen maatregel. Er zijn verschillende alternatieve verklaringmogelijkheden voor het effect:

1. er is een algemeen dalende trend (daar is bij verkeersslachtoffers inderdaad sprake van).
 2. er is sprake van regressie naar het gemiddelde. Maatregelen worden doorgaans genomen in gevaarlijke situaties. In dat geval kan alleen al op grond van toeval in de nameting het aantal richting gemiddelde opschuiven.
 3. verandering in registratie (de registratie van verkeersslachtoffers verslechtert).
 4. andere externe gebeurtenissen hebben de situatie veranderd, en dat verklaart het effect. Bijvoorbeeld: tegelijkertijd met de herinrichting van kruispunten is de maximumsnelheid verlaagd. Wordt een verlaging van het aantal slachtoffers nu teweeg gebracht door de kruispuntherinrichting, de verlaagde maximumsnelheid of juist de combinatie van beide?
 5. een combinatie van de mogelijkheden 1 t/m 4.
- Voor nadere uitwerking, zie bijlage 1, par. 4.

5.3 De praktijk – observationeel onderzoek

Meestal is het in verkeersveiligheidsonderzoek om praktische of ethische redenen niet mogelijk om onderzoeksobjecten random te selecteren uit de doelpopulatie, dan wel om onderzoeksobjecten random toe te wijzen aan een experimentele en een controlegroep. Een voorbeeld is de onderzoeksvraag of autobestuurders onder invloed van alcohol, drugs of medicijnen (of een combinatie hiervan) een verhoogd risico hebben om bij een ongeval betrokken te raken. In dit geval is het niet mogelijk om een aselechte steekproef van autobestuurders random toe te wijzen aan de experimentele en de controlegroep, en de bestuurders in de experimentele groep vervolgens onder invloed van alcohol, drugs of medicijnen de weg op te sturen. Meer in het algemeen is het om ethische redenen niet mogelijk om proefpersonen random toe te wijzen aan omstandigheden die kunnen resulteren in menselijk lijden.

Veel onderzoeken naar de effecten van verkeersveiligheidsmaatregelen zijn om die redenen dan ook niet gebaseerd op experimentele onderzoeksdesigns maar puur observationeel van aard. Het verloop van de genomen maatregel(en) wordt gevolgd. In tegenstelling tot de ideale experimentele opzet (zie par. 5.2) kunnen echter geen harde oorzaak-gevolg conclusies aan de observationele studies worden verbonden.

Desondanks is het de kunst om ook in deze situatie te proberen de ideale onderzoeksopzet zoveel mogelijk te benaderen door:

1. te proberen vast te stellen wat de verkeersveiligheidssituatie zou zijn geweest indien de maatregel niet was ingevoerd;
2. statistisch in plaats van experimenteel te controleren voor versturende variabelen;
3. de (steekproef uit de) doelpopulatie te vergelijken met een (steekproef uit een) referentiepopulatie waarop de onderzochte maatregel geacht wordt geen effect te hebben gesorteerd, maar die verder zoveel mogelijk dezelfde kenmerken heeft als de doelpopulatie.

We bespreken twee methoden waarmee deze drie manieren om de ideale onderzoeksopzet te benaderen kunnen worden gerealiseerd, te weten tijdreeksanalyse en de zogenoemde Empirical Bayes methode.

Tijdreeksanalyse

De eerste methode is de tijdreeksanalyse. Een van de taken van verkeersveiligheidsonderzoek is het volgen, beschrijven, verklaren en voorspellen van ontwikkelingen op het terrein van de verkeersveiligheid. Als deze ontwikkelingen worden gekwantificeerd, dan zijn er analysetechnieken nodig om op verantwoorde wijze ontwikkelingen te kunnen volgen, beschrijven, verklaren of voorspellen. De term 'ontwikkelingen' impliceert dat het gaat om de analyse van een speciaal type gegevens. Het betreft gegevens die bestaan uit herhaalde metingen over tijd van een bepaald aspect van het verkeersproces. Een voorbeeld zijn de jaarlijkse aantallen verkeersdoden geobserveerd over een bepaalde periode. De reeks meetgegevens van zo'n variabele wordt een tijdreeks genoemd. Een belangrijke eigenschap van de observaties in een tijdreeks is dat ze meestal niet onafhankelijk zijn van elkaar: het geobserveerd aantal verkeersdoden van vorig jaar is immers meestal een vrij goede indicator voor het aantal verkeersdoden van dit jaar. Aangezien standaardtechnieken uitgaan van onafhankelijke observaties, resulteert de analyse van tijdreeksen met standaardtechnieken heel vaak in residuen die eveneens onderling gecorreleerd zijn. Dit terwijl statistische toetsen en betrouwbaarheidslimieten gebaseerd zijn op de cruciale aanname dat de gevonden residuen random zijn, en dus onafhankelijk van elkaar (SWOV factsheet sept 2010).

De analyse van tijdreeksen vereist speciale modellen die in staat zijn om rekening te houden met de afhankelijkheden in de observaties. Een van deze modellen zijn de zogenoemde state space modellen (zie Harvey, 1989, Durbin en Koopman, 2001, Commandeur en Koopman, 2007). Naast het uitvoeren van relatief simpele beschrijvende analyses kunnen deze modellen ook gebruikt worden om:

- de effecten van maatregelen en van andere variabelen op de ontwikkeling in de verkeersveiligheid te evalueren;
- te onderzoeken of nieuw gepubliceerde cijfers al dan niet afwijken van de verwachting op grond van het verleden;
- prognoses te doen over toekomstige ontwikkelingen in de verkeersveiligheid.

(SWOV factsheet sept 2010). De uitwerking van tijdreeksanalyse modellen voor het ex post evalueren van effecten van verkeersveiligheidsmaatregelen in observationeel onderzoek is te vinden in bijlage 1, par. 6, alwaar o.a. wordt besproken hoe deze modellen kunnen worden gebruikt om een inschatting te verkrijgen van de ontwikkeling in het aantal verkeersslachtoffers of -ongevallen als de maatregel niet was ingevoerd, om het effect van de maatregel te bepalen door middel van een interventie analyse, om bij zo'n interventie analyse expliciet te

corrigeren voor versturende variabelen door ze als covariaten in de analyse mee te nemen, en om voor niet gemeten versturende variabelen te corrigeren door een referentiegroep bij de analyse te betrekken.

Empirical Bayes

De tweede methode is de Empirical Bayes (EB) methode. De hier beschreven aanpak is gebaseerd op Hauer (1997). In de EB methode wordt veel aandacht besteed aan het feit dat de observatie-eenheden waarop de onderzochte maatregel wordt toegepast in observatiestudies vaak geen aselechte steekproef uit de doelpopulatie vormen maar juist voor behandeling gekozen worden *omdat ze op dat moment relatief erg onveilig zijn (in termen van aantallen ongevallen en/of slachtoffers)*. Aangezien er dan een duidelijke relatie bestaat tussen het aantal ongevallen en/of slachtoffers van de studie objecten en de reden dat juist deze studie objecten aan de maatregel worden onderworpen is er sprake van *selectie bias* hetgeen aanleiding geeft tot het *regressie naar het gemiddelde effect*. Daarnaast benadrukt Hauer dat in een naïeve voor- en nastudie het gevonden effect niet alleen het gevolg hoeft te zijn van de onderzochte maatregel, maar ook van versturende variabelen als verkeersintensiteit, het weer, de omvang van het voertuigpark, het gedrag van bestuurders, de registratiegraad van ongevallen, etc., etc.

De EB methode voor de ex post evaluatie van de effecten van verkeersveiligheidsmaatregelen bestaat uit de volgende stappen:

1. observeer het aantal ongevallen voor de eenheden waarop de maatregel is toegepast (of een steekproef daaruit) in de voor- en naperiode;
2. bepaal het verwachte aantal ongevallen voor diezelfde eenheden in de voorperiode door het aantal ongevallen en de variatie daarin in een referentie populatie te bepalen;
3. corrigeer het geobserveerde aantal ongevallen in de voorperiode voor het verwachte aantal ongevallen in de referentie groep;
4. vergelijk dit laatste gecorrigeerde aantal ongevallen uit de voorperiode met het geobserveerde aantal ongevallen in de naperiode, en baseer op *dit verschil* het effect van de maatregel.

Met deze aanpak wordt getracht zoveel mogelijk te corrigeren voor het regressie naar het gemiddelde effect dat optreedt wanneer bij een observationele studie het vermoeden bestaat dat er een samenhang is tussen de maatregel zelf en de reden waarom de maatregel juist bij deze onderzoekseenheden wordt onderzocht en toegepast. Voorwaarde voor de toepasbaarheid van de EB methode is uiteraard dat er schattingen beschikbaar moeten zijn over de verwachte waarde van het studie object middels onderzoek naar de samenhang tussen ongevallen en/of slachtoffers en de eigenschappen van objecten in een referentie populatie. De uitwerking van de Empirical Bayes methode is te vinden in bijlage 1, par 7.

5.4 Conclusies

Duidelijk is dat de bewijsvoering rond de ex post evaluatie van verkeersveiligheidsmaatregelen in een natuurlijke setting nooit helemaal waterdicht zal zijn, met name in termen van causaliteit. Observationele studies kunnen echter aanzienlijk sluitender en overtuigender worden gemaakt door - waar mogelijk:

1. langere tijdreeksen van observaties in de voorperiode te gebruiken om te kunnen corrigeren voor algemene trends en regressie naar het gemiddelde effecten;

2. voor verstorende variabelen die bekend én tijdens voor- en nameting gemeten zijn expliciet te corrigeren door ze als covariaten mee te nemen in de interventie analyse;
3. voor verstorende variabelen die onbekend en/of niet gemeten zijn impliciet te corrigeren door een referentiegroep bij de analyse te betrekken;
4. verschillende soorten analyse te gebruiken om effectschattingen van een maatregel te kunnen kruisvalideren.

In de complexe praktijk spelen vele variabelen een rol. De ideale experimentele onderzoeksopzet is bij de effectevaluatie van beleidsmaatregelen in het verkeer zelden of nooit realiseerbaar. Ook kan niet altijd worden beschikt over betrouwbare of goed geregistreerde ongevalldata. In dat geval kunnen data worden verzameld over indicatoren die met ongevallen verband houden, zoals conflicten, snelheden of bepaalde soorten bekeuringen. Afhankelijk van de beschikbaarheid van data kan uiteindelijk worden gekozen voor een kwantitatieve of kwalitatieve methode van evalueren, of een combinatie van beide. Door beide methodes te combineren kunnen in veel gevallen uitspraken over de werking van maatregelen beter worden onderbouwd. Het volgende hoofdstuk gaat hier verder op in.

6. Beschouwing: wat is zinvol en haalbaar?

6.1 complexiteit van de problematiek

Het is duidelijk dat bij het opzetten van beleidsevaluaties veel vragen moeten worden beantwoord. Met de vaak grote complexiteit van problemen bij beleidsevaluatie kan rekening worden gehouden, door:

- beleidsdoelen zo helder en concreet mogelijk te formuleren;
- algemene beleidsdoelen coherent door te vertalen naar beleidsvoornemens, beleidsplannen en uitvoeringsmethoden en –middelen;
- zicht te houden op de vele instanties en stakeholders die betrokken zijn bij beleidsvoorbereiding en/of –uitvoering, en ze ook te betrekken bij de evaluatie;
- verstandig om te gaan met de vele parallelle uitvoeringsactiviteiten die elkaar onderling beïnvloeden;
- wijzigingen die gaandeweg de uitvoering worden aangebracht goed te documenteren, zodat bij evaluatie daarvoor kan worden gecorrigeerd;
- de indicatoren en de criteria in overleg zo duidelijk of eenduidig mogelijk vast te stellen;
- verschillende soorten indicatoren op verstandige wijze met elkaar in verband te brengen, waarbij een goede afweging wordt gemaakt tussen de toegekende soorten “waarden” van elke indicator. Sommige indicatoren zijn goed in geld uit te drukken, andere representeren emotionele of behoefegerelateerde waarden;
- omdat het zelden mogelijk is om een ideaal experiment op te zetten (zie hoofdstuk 5), moet steeds worden overwogen wat in de context van de beleidsvraag, de beleidsruimte en het beleidsonderwerp “next best” is;
- omdat er vaak vele alternatieve verklaringen voor gevonden of uitgebleven effecten zijn te bedenken, die niet gemakkelijk toetsbaar zijn is het nodig om vanuit verschillende referentiekaders naar aanpak, meting en resultaat te kijken;
- datzelfde is van belang om de mix van bedoelde en onbedoelde en gewenste en ongewenste prestaties of effecten te ontwarren, en daarover een gewogen oordeel te geven;
- te waarborgen dat de relatie tussen ex ante, ex durante en ex post zo consistent mogelijk wordt gelegd;
- deskundigen met uiteenlopende visie op de functie, aanpak en de verwachtingen van evaluaties bij elkaar te brengen;
- zorg te dragen voor goede informatiebronnen voor het uitvoeren van evaluaties, wat betekent dat gedegen documentatie bewaard moet blijven over de beleidsvoornemens, de procedures, de uitvoering, en de besluitvorming rond een project of maatregel.

Bij verkeersveiligheid moet ook rekening gehouden worden met specifieke problemen, door:

- bij data verzameling en analyse er rekenschap van te geven dat veel maatregelen niet op één moment worden geïmplementeerd, maar gedurende een aantal jaren. Bij analyse moet dan rekening worden gehouden met tijds- en interactie-effecten;
- bij elkaar overlappende maatregelen trachten te corrigeren voor de interactie-effecten (indien men de werking van een bepaalde maatregel wil kennen). Dit kan in een aantal situaties door een wetenschappelijke methode te gebruiken (zie hoofdstuk 5), maar vaak zal men op basis van theorie of ervaringen inschattingen

maken van de relatieve bijdrage van verschillende maatregelen aan het totale effect.

- bij kleine effecten die niet altijd significant aantoonbaar zijn, na te gaan in hoeverre maatregelen met een klein effect een wenselijke bijdrage leveren aan andere maatregelen, andere maatschappelijke doelen of wenselijke processen;
- rekening te houden met het feit dat de verkeersongevallen dataset van de politie niet compleet, niet stabiel in de tijd en niet representatief is. De gegevens kunnen toch worden gebruikt door correcties toe te passen, of door zinvolle aanvullingen uit andere bronnen toe te voegen;
- rekening te houden met het gegeven dat niet alle problemen of variabelen even gemakkelijk zichtbaar zijn te maken (naar Rumar, 1999). Er zijn direct meetbare variabelen zoals ongevallen of snelheid (eerste orde), variabelen die een diepgaandere analyse vereisen, zoals regelgeving of genoten opleiding van weggebruikers (tweede orde) en variabelen die bijna geheel verborgen blijven, zoals probleembewustzijn of de invloed van managementsystemen (derde orde). Het kost veel investering om de tweede en derde orde variabelen meetbaar te maken en te meten. Er moet worden afgewogen wat zinvol is en de moeite waard om te investeren.

6.2 Afwegingen

Gezien de complexiteit van ex post evaluatie moeten afwegingen worden gemaakt over wat en hoe evalueren in de beleidscontext.

Met betrekking tot het *wat* gaat het om de vraag of beleid het effect van specifieke maatregelen wil weten, wil weten door welke factoren het verloop van verkeersslachtoffers of van ongevallen over een bepaalde periode wordt bepaald, of wil weten wat er van bepaalde ontwikkelingen kan worden geleerd.

Al deze vragen worden gesteld, en de beantwoording vereist een overwogen aanpak. Met de vraag naar het *wat* gaat de vraag over het *hoe* samen.

Bij de vraag naar het effect van een specifieke maatregel kan worden nagegaan of een van de in hoofdstuk 5 genoemde aanpakken, zo nodig aangevuld met kwalitatief onderzoek (zie par. 3.3.) kan worden gehanteerd om met voldoende mate van plausibiliteit uitspraken te doen over de werking van een maatregel. Dat geeft dan voeding aan de beleidsoverweging om een specifieke maatregel te continueren, uit te breiden, te wijzigen, of eventueel af te schaffen.

Bij het verklaren van het verloop van verkeersonveiligheid moet het hele pakket aan ontwikkelingen en maatregelen en de bijdrage ervan aan verkeersonveiligheid over een bepaalde periode worden beoordeeld. Geprobeerd kan worden om in een multifactorieel verklaringsmodel met vele afhankelijke en onafhankelijke variabelen na te gaan welke variabelen met elkaar samenhangen, en hoe die samenhang kwantitatief en kwalitatief kan worden geded. Zo'n verklaringsmodel is niet deterministisch en kan niet uitputtend zijn. Omdat bij het verloop van ongevallen of slachtoffers zowel autonome ontwikkelingen, als een pakket van maatregelen, als toeval alle een rol spelen, kan ook naderhand niet alles worden verklaard. Daarom moet een afweging worden gemaakt tussen de vereiste inspanning en kosten om zoveel mogelijk te verklaren, en wat bij een bepaalde mate van inspanning en kosten voldoende verklaring oplevert voor beleidsbesluiten.

Om de betrouwbaarheid en validiteit van de resultaten te vergroten kan gebruik worden gemaakt van triangulatie (zie hoofdstuk 3, par. 3.3).

De vraag naar het leren is belangrijk als een project of pakket innovatieve elementen bevat, die bij vervolgtoeepassingen of andere toepassingen worden gehanteerd. De ervaringen van verschillende betrokkenen moeten goed worden gedocumenteerd, om van daaruit het bestaande kennisbestand te verrijken.

Uit voorstaande beschouwing zijn verschillende specifieke afwegingen te destilleren (zie ook Berveling et al, 2009):

- moeten alle projecten worden geëvalueerd, of kan met een beperkt aantal worden volstaan. Weegt de investering op tegen de kosten, en kunnen op basis van een beperkt aantal evaluaties voldoende conclusies worden getrokken, of kan genoeg worden geleerd?
- moeten alle aspecten van een maatregelenpakket op het terrein van verkeersveiligheid in de evaluatie in even grote mate worden meegenomen? Dit hangt af van wat er al aan kennis is over bepaalde aspecten, hoe zwaarwegend het aspect is in het geheel van maatregelen, en hoe onderzoekbaar bepaalde aspecten zijn.
- wanneer moet of kan met een lichte evaluatie worden volstaan? In het algemeen geldt dat met een lichte evaluatie kan worden volstaan als beleid diepgaande studies niet omarmt, als er tijdsdruk is om tot een advies te komen, als de bestaande of toegekende middelen ontoereikend zijn, als er te weinig informatie bij de uitvoering van een maatregelenpakket beschikbaar komt, als de speelruimte voor beleid gering is, als wordt voorzien dat maatregelen op afzienbare termijn weer door andere worden vervangen, als er uit andere analoge projecten al veel bekend is, of als een maatregel uit haar aard weinig invloed op verkeersveiligheid kan hebben.
- op welke momenten moet worden geëvalueerd? Hier kan een onderscheid worden gemaakt in procesevaluatie en eindevaluatie. Afhankelijk van looptijd en diepgang van een maatregelenpakket kan tijdens het proces een tot twee keer een evaluatie (ex durante) plaatsvinden. Dit zal doorgaans een lichte evaluatievorm zijn. Na afloop kan eveneens vaker geëvalueerd worden, waarbij rekening moet worden gehouden met enerzijds eventuele ingroei-effecten, anderzijds verstoringen op langere termijn door nieuwe ontwikkelingen. Berveling et al (2009) opteren daarom voor een "zwarte" ex post evaluatie na 3 jaar.
- in welke mate moet elke evaluatie wetenschappelijk verantwoord zijn? Dat hangt enerzijds samen met het doel van evalueren, anderzijds met beschikbare middelen of instrumenten. Er zijn maatregelen die lastig wetenschappelijk te onderzoeken zijn, zoals flankerend beleid of het effect van een aantal nieuwe verkeersregels; dus tweede en derde orde variabelen. Dat komt omdat ze deel uitmaken van een complex aan maatregelen, maar ook omdat hun effect vaak niet direct te meten is. Er zal dan gebruik gemaakt moeten worden van instrumenten zoals opiniepeilingen, kennismetingen, of expertbeoordelingen. Als die allemaal dezelfde richting opwijzen, dan is het plausibel om aan te nemen dat er sprake is van een bepaald effect (of geen effect). Ook als er te weinig middelen beschikbaar zijn voor een uitgebreide wetenschappelijke evaluatie zullen concessies aan de wetenschappelijkheid gedaan moeten worden.

6.3 zinvol ex post evalueren

Idealiter wordt voor ex post beleidsevaluatie een evaluatieprogramma opgesteld, met in elk geval de volgende elementen:

- wat is de probleemstelling; op welke vraag of vragen moet onderzoek antwoord geven, en wanneer?,
- welke doelen moeten worden geëvalueerd (kan een verplichtend karakter hebben, zoals in het kader van een wet of verdrag, of toezeggingen aan de Kamer), zijn de doelen duidelijk beschreven en goed afgebakend?,
- welke onderzoeksvragen kunnen uit probleemstelling en doelformulering worden afgeleid, zijn daarvoor hypothesen op te stellen?,
- welke indicatoren en criteria worden gehanteerd?,
- welke methodes en meetinstrumenten worden voor ex post evaluatie gehanteerd?,
- hoe wordt de ex post evaluatie georganiseerd, van welke informatiebronnen wordt gebruik gemaakt?,
- wanneer worden de verschillende momenten voor ex post gepland?,
- welke partijen zijn bij de evaluatie betrokken en wat is hun rol en functie?,
- hoe worden de uitkomsten van de ex post evaluatie weergegeven en gepresenteerd?,
- welke verklaringen, voorspellingen en aanbevelingen worden gegeven?,
- is te verwachten dat op basis van de uitkomsten beleid wordt gewijzigd of geleerd wordt van de evaluatie?,
- hoeveel tijd, inspanning, geld vergt de evaluatie?.

In veel gevallen is het mogelijk om richting te geven aan de bovengenoemde elementen voor een programma van evaluatie. Dat wil niet zeggen dat elk van de hierboven gestelde vragen een ideaal antwoord heeft dat leidt tot de meest optimale aanpak van een ex post evaluatie. De kunst van het evalueren is telkens bekijken wat onder de gegeven omstandigheden mogelijk is. Als er een punt wordt bereikt waarbij niet op doelmatige wijze een ex post evaluatie van voldoende kwaliteit kan worden geleverd, dan zal dit in elk geval moeten worden gemeld.

Op basis van de in dit hoofdstuk gegeven beschouwing kan worden geconcludeerd dat:

- gezien de complexiteit en context van beleidsevaluatie voor het besluit over de aard en diepgang van ex post evaluatie verschillende vragen moeten worden beantwoord.
- de beantwoording van deze vragen moet leiden tot afwegingen over wat, hoe en wanneer evalueren.
- overwegingen voor ex post evaluatie tijdig moeten worden gemaakt, zodat hier bij ex ante en ex durante evaluatie rekening mee gehouden kan worden.
- op doelmatige wijze de voor beleid benodigde informatie moet worden verzameld, waarvan de kwaliteit moet worden aangegeven.
- per geval moet worden bekeken of ex post evaluatie noodzakelijk is, welke eisen daaraan gesteld moeten worden in het licht van de beleidsvragen, de beleidscontext, en wat mogelijk is gezien de beschikbare informatie, middelen en tijd.

In hoofdstuk 7 wordt daarvoor een voorstel gedaan.

7. Conclusies

7.1 Conclusies

In de inleiding (par. 2.4, blz. 9 en 10) zijn een zestal onderzoeksvragen gesteld. Op grond van deze literatuurstudie is geprobeerd een zo goed mogelijk antwoord op de vragen te geven.

Algemeen kan worden geconcludeerd dat bij beleidsbesluiten over verkeersveiligheid doorgaans wel een evaluatie ex ante wordt uitgevoerd, maar dat niet altijd voldoende wordt nagedacht over ex post evaluatie. Daardoor blijft de verantwoording en het leren van beleid onderbelicht. Het is nodig om tijdig een keuze te maken over wat wel en niet ex post geëvalueerd moet worden om essentiële beleidsvragen te beantwoorden of om te leren voor vervolgbeleid of vervolgonderzoek. De keuzes voor ex post evaluatie moeten liefst in de fase van beleidsbesluitvorming worden gemaakt, zodat vanaf de ex ante fase hiermee rekening kan worden gehouden. Vervolgens kunnen de vragen over hoe, wat en wanneer evalueren worden opgenomen in een evaluatieplan, waarin mogelijk verschillende varianten van evaluatie en de gevolgen voor in te zetten menskracht en middelen en het tijdspad worden meegenomen. Op basis daarvan kunnen weloverwogen definitieve keuzes worden gemaakt over wat te evalueren en met welke diepgang.

Hierna wordt ingegaan op de in de inleiding gestelde vragen.

Vraag 1. Hoe kunnen we 'ex post' evaluatie definiëren?

Ex post evaluatie is in deze studie breed gedefinieerd als systematisch onderzoek achteraf naar de uitvoering, realisatie, doelmatigheid en effectiviteit van verkeersveiligheidsbeleid. Gedegen evaluaties maken idealiter onderdeel uit van een cyclus van besluitvorming. Bij besluitafwegingen geeft ex ante evaluatie aan welke richtingen mogelijk zijn. Na besluitneming geven ex durante en ex post evaluaties inzicht in de realisatie en de effecten. Goede ex post evaluatie vereist goede ex ante evaluatie, en goede ex ante evaluatie vergt weer goede ex post evaluatie.

Vraag 2. Wat mogen we wel en niet van een ex post evaluatie verwachten? Wanneer is het zinvol om een ex post evaluatie uit te voeren? Wat zouden de doelen van een ex post evaluatie kunnen zijn?

Een ex post evaluatie in ideale omstandigheden kan antwoord geven op de vraag of een maatregel of beleidspakket verandering in aantal en/of aard van verkeersongevallen of slachtoffers heeft *veroorzaakt*. Een ideaal design kan worden beschreven. Omdat verkeersveiligheidsmaatregelen in een dynamische en complexe omgeving worden genomen is dit ideale design in de praktijk niet bruikbaar. Verkeersveiligheidsbeleid en de daaruit voortvloeiende maatregelen behelzen doorgaans meerdere aspecten, zoals verandering van wetgeving, technische veranderingen en flankerende maatregelen. Daarom is het lastig om de precieze werking van elk onderdeel te isoleren en om de causaliteit ondubbelzinnig vast te stellen.

Er zijn verschillende benaderingswijzen van evaluatie. Afhankelijk van de beleidsvraagstelling en van de complexiteit van de context waarin maatregelen worden genomen, kan voor een benaderingswijze (evidence based, realistic, complexity, onderhandelend evalueren) worden gekozen.

Ex post evaluatie kan de functie hebben van verantwoorden en van leren (of beide). De verantwoordingsfunctie vraagt uitspraken over de werking van uitgevoerd beleid,

zowel naar uitvoerbaarheid, efficiëntie als effect. De leerfunctie vraagt uitspraken over het verloop van uitvoering, de gekozen aanpak en gebruik van methoden en technieken. Elk van deze aspecten kan het uitvoeren van een ex post evaluatie waardevol maken. Daarbij is het nodig om een afweging te maken tussen investering voor ex post evaluatie en verwachte opbrengst.

Vraag 3. Hoe zou een methode die de besparing van slachtoffers en eventueel de kosteneffectiviteit achteraf kan evalueren (i.e. een ex post evaluatie) eruit kunnen zien?

Het ideale design met aselecte steekproeftrekking, voor- en nameting, controle over variabelen is zoals gezegd in de praktijk nauwelijks of niet haalbaar. Er zal rekening gehouden moeten worden met minder ideale situaties, waarbij door middel van andere methoden uitspraken over de effecten kunnen worden gedaan. Daarbij is het belangrijk om langere tijdreeksen te gebruiken, waar mogelijk expliciet (gebruik als covariaten) of impliciet (gebruik referentiegroep) te corrigeren voor versturende variabelen en verschillende soorten analyse te hanteren.

Het is mogelijk om plausibiliteitsuitspraken te doen op basis van lichtere vormen van evaluatie die methodologisch niet aan alle eisen voldoen, maar waarbij toch met enige mate van zekerheid conclusies kunnen worden getrokken. Door een mix van onderzoeksaanpakken, informatiebronnen en beoordelaars te gebruiken kan de plausibiliteit worden vergroot.

Vraag 4. In welke gevallen is het mogelijk om een zinvolle ex post evaluatie uit te voeren?

Dit is in het algemeen lastig te zeggen. Het zal van geval tot geval moeten worden beoordeeld. Wel kan worden gezegd dat procesevaluatie op kortere termijn kan worden gerealiseerd dan productevaluatie of evaluatie van outcome. Als op korte termijn moet worden geëvalueerd of met weinig middelen, dan zal de uitkomst grofmazig en doorgaans slechts indicatief zijn. Het vaststellen van effectiviteit vereist meer tijd en middelen.

Ook is geconcludeerd dat een zinvolle ex post evaluatie niet altijd mogelijk is of dat een evaluatie niet altijd nodig is. Daarbij kan worden gedacht aan situaties zoals het ontbreken van een goede en duidelijke beleidsvraag, het ontbreken van indicatoren of data, het ontbreken van middelen om een evaluatie goed uit te voeren, het stellen van een onrealistisch tijdstip waarop evaluatie moet zijn uitgevoerd, of het ingevoerd hebben van een maatregel waarvan de aanpak en effecten al door en door bekend zijn.

Vraag 5. Wanneer wordt voor welk type en welke methode van ex post evaluatie gekozen?

Het opzetten van een zuiver experiment, waarbij variabelen eenduidig worden benoemd en gemeten vereist voorwaarden waaraan zelden wordt voldaan. In sommige situaties kunnen alternatieve opzetten, zoals een experiment zonder voormeting of een experiment zonder controlegroep worden gehanteerd.

Als het niet mogelijk is om random te selecteren of toe te wijzen aan experimentele en controlegroep, wat bij verkeersveiligheidsmaatregelen meestal het geval is, dan wordt gekozen voor observationeel onderzoek. Dat is onderzoek waarbij het mogelijke effect van de maatregel wordt bepaald door het aantal slachtoffers of ongevallen vóór invoering van de maatregel te vergelijken met het aantal na invoering van de maatregel, zonder de omstandigheden bij invoering van de maatregel experimenteel te manipuleren. Als over langere periodes data worden

verzameld kan gebruik worden gemaakt van tijdreeks analyse, die robuuster wordt door gebruik te maken van toevoeging van versturende variabelen (indien gekend en gemeten), of vergelijking met een referentiegroep.

Een andere mogelijkheid is het gebruik van de Empirical Bayes methode, waarbij wordt gecorrigeerd voor het regressie naar het gemiddelde effect. Dit gebeurt door gebruik te maken van gegevens in een referentiegroep, de gegevens in de geobserveerde groep in de voorperiode daarvoor te corrigeren, en dit resultaat te vergelijken met de gegevens in de naperiode.

Voor de toepassing van de methodes zijn zowel verkeersonveiligheidsdata (slachtoffers en ongevallen) als data over mate van invoering van het beleid, als veranderingen op basis van ingevoerd beleid (denk aan gedragsveranderingen zoals snelheidsgedrag, gordeldracht of alcoholgebruik) nodig.

Als niet kan worden beschikt over betrouwbare ongevallendata kunnen data worden verzameld over gedragingen die met ongevallen verband houden, zoals conflicten, snelheden of bepaalde soorten bekeuringen. Afhankelijk van de aard en beschikbaarheid van data, de vereiste diepgang van analyse en de beleidscontext kan worden gekozen voor een kwantitatieve of kwalitatieve methode van evalueren. Door middel van triangulatie kan de betrouwbaarheid en validiteit van data en daarop gebaseerde uitspraken worden vergroot.

Kwalitatieve en kwantitatieve methoden kunnen elkaar goed aanvullen.

Vaak geldt dat voor "losse maatregelen" een andere opzet wordt gebruikt dan voor het beoordelen van het totale beleidspakket. Voor het laatste kan een analysemodel met verschillende variabelen en hun relaties worden gehanteerd.

Die worden, afhankelijk van vraagstelling en beleidscontext en detail geanalyseerd, maar kunnen ook globaler als één pakket worden geëvalueerd.

Vraag 6. Welke conclusies kunnen worden getrokken?

Deze vraag is in het eerste deel van dit hoofdstuk beantwoord (blz. 32).

Literatuur

Algemene Rekenkamer (2005). *Handleiding onderzoek naar doelmatigheid en doeltreffendheid*. Den Haag.

Ampe, J., T. Geudens en C. Macharis (2008). *Multi-criteria Analyse en Multi-Actor Multi-Criteria Analyse*. Steunpunt mobiliteit en openbare werken, Hasselt.

Beenker, N. & Verdiesen, J. (2010). *Bouwstenen voor een veiligheidsambitie: onderzoek naar verkeersveiligheidseffecten van maatregelscenario's gericht op rijkswegen*. DVS, Delft.

Berveling, J., W. Groot, M. Lijesen, F. Savelberg, E. van der Werff (2009). *Na het knippen van het lint. Het ex post evalueren van infrastructuur*. KiM, Den Haag.

Commandeur, J.J.F. & Koopman, S.J. (2007). *An introduction to state space time series analysis*. Oxford University Press, Oxford.

COST 329 (2004). *Models for traffic and safety developments and interventions* (Final report of the action EUR 20913). Directorate General for Transport, European Commission, Brussel.

CROW (2008). *Handboek verkeersonderzoek*. CROW-publikatie 248. CROW, Ede.

CROW (2008). *Handboek verkeersveiligheid*. CROW-publikatie 261. CROW, Ede.

Durbin, J. & Koopman, S.J. (2001). *Time Series Analysis by State Space Methods*. *Oxford statistical science series*, nr. 24. Oxford University Press, Oxford.

Elvik, R. & Vaa, T. (2004). *The handbook of road safety measures*. Emerald Group Publishing Limited, Bingley.

Hagenzieker, M.P. (1995). *Bromfietzers op de rijbaan: ongevalstudie*. R-95-33. Stichting Wetenschappelijk Onderzoek Verkeersveiligheid SWOV, Leidschendam.

Hanemaayer, D. (2010). *Ex post evaluatie ten behoeve van verkeersveiligheidsbeleid*. DVS, Delft.

Harvey, A.C. (1989). *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press, Cambridge.

Harvey, A.C. & Durbin, J. (1986). The effects of seat belt legislation on British road casualties: A case study in structural time series modelling. *Journal of the Royal Statistical Society A*, vol. 149, nr. 3, p. 187-27.

Hauer, E. (1997). *Observational Before-After Studies in Road Safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety*. Pergamon, New York.

Hoff, J.P. van 't, Nägele, R.C. & Zoutendijk, D.C. (2000). *Evaluatiekader BER-Verkeersveiligheid: Eindrapport*. Traffic Test, Veenendaal.

Kanji, G.K. (1993). *100 statistical tests*. Londen: Sage Publications.

Kirk, R.E. (1968). *Experimental design: Procedures for the behavioral sciences*. Wadsworth Publishing Company, Belmont.

Langeveld, P.M.M. & Schoon, C.V. (2004). *Kosten-batenanalyse van maatregelen voor vrachtauto's en bedrijven*. R-2004-11. Stichting Wetenschappelijk Onderzoek Verkeersveiligheid SWOV, Leidschendam.

Loon, A. van (2001). *Evaluatie verkeersveiligheidseffecten "Bromfiets op de rijbaan"*. AVV, Rotterdam.

Morsink, P.L.J. & Wismans, L.J.J. (2008). *Verkeersmodellen en verkeersveiligheid*. Stichting Wetenschappelijk Onderzoek Verkeersveiligheid SWOV, Leidschendam.

Peuter, B. de, J. de Smedt, en G. Bouckaert (2007). *Handleiding beleidsevaluatie. Deel 1: Evaluatiedesign en -management*. Leuven.

Peuter, B. de, J. de Smedt, W. van Dooren en G. Bouckaert (2007). *Handleiding beleidsevaluatie. Deel 2: Monitoring van beleid*. Leuven.

Rumar, K. (1999). *Transport Safety Visions, Targets and Strategies: Beyond 2000*. Brussel: ETSC.

Wouters, P, H. Derriks en H. van der Loop (2009). *Toezicht tellen*. KiM, Den Haag.

Twisk, D.A.M., Vlakveld, W.P. & Commandeur, J.J.F. (2007). *Wanneer is educatie effectief?* R-2006-28. Stichting Wetenschappelijk Onderzoek Verkeersveiligheid SWOV, Leidschendam.

Vlist, A.J. van der; Bunte, F.H.J en Galen, M.A. van (2007). *Beleidsvaluatie ex-post. Methodiek en illustratie*. LEI, Den Haag.

Waal, A.A. de en Kerklaan, L.A.F.M. (2004). *De Resultaatgerichte Overheid. Op weg naar de prestatiegedreven overheidsorganisatie*. SDU, Den Haag.

Wilmots, B. (2009). *Verkeersveiligheidsonderzoek*. Imob, Universiteit Hasselt.

Zee, F. van der (2004). *Kennisverwerving in de empirische wetenschappen, de methodologie van wetenschappelijk onderzoek*. BMOOO, Groningen.

Bijlage 1: Methodologische uitwerkingen

In deze bijlage laten we zien hoe de ex post effect evaluatie van een verkeersveiligheidsmaatregel er in het ideale geval zou moeten uitzien, willen we na afloop daadwerkelijk kunnen concluderen dat de maatregel een bepaald effect heeft *veroorzaakt*. Kort gezegd komt het erop neer dat een ideale effect evaluatie alleen kan worden uitgevoerd als we in staat zijn om de omstandigheden tijdens de invoering van een maatregel te *manipuleren*. Zoals we in paragraaf 5 e.v. zullen bespreken is het bij de evaluatie van een verkeersveiligheidsmaatregel echter meestal onmogelijk om op een experimentele manier invloed uit te oefenen op de omstandigheden waaronder de maatregel wordt ingevoerd.

De reden dat we in deze bijlage toch eerst uitgebreid ingaan op de 'ideale' onderzoeksopzet is dat de filosofie en principes die eraan ten grondslag liggen ons veel kunnen leren over de beste manier om het effect van een maatregel te evalueren na invoering in een natuurlijke observationele setting.

We lichten de ideale ex post evaluatie toe aan de hand van de invoering van de volgende infrastructurele maatregel: het ombouwen van kruispunten naar rotondes.

1. De onderzoeksvraag, en de operationalisatie van variabelen

Het opzetten van een experimentele ex post evaluatie begint net als alle onderzoek met het formuleren van een *onderzoeksvraag*. Bij de ex post evaluatie van het ombouwen van kruispunten naar rotondes zou de onderzoeksvraag kunnen luiden:

Welk effect heeft het ombouwen van kruispunten naar rotondes op de verkeersveiligheid?

Het eerste dat over deze onderzoeksvraag kan worden opgemerkt is dat zij zo vaag en algemeen is opgesteld dat er geen concreet onderzoek mee kan worden uitgevoerd. Deze onderzoeksvraag roept namelijk zelf weer de volgende belangrijke vraag op: wat wordt er bedoeld met 'de verkeersveiligheid'? Is dat het aantal ongevallen dat plaatsvindt op kruispunten en rotondes (ongeacht de afloop)? Of is dat het aantal ongevallen met dodelijke afloop? Of het aantal doden? En zijn dat aantallen zoals geregistreerd door de politie? Of door verzekeringsmaatschappijen? Of door ziekenhuizen? Het antwoord op deze vragen zorgt ervoor dat een relatief abstract begrip als 'de verkeersveiligheid' zodanig concreet wordt gemaakt dat het ook echt kan worden gemeten. Deze vertaalslag wordt in onderzoek ook wel het **operationaliseren van variabelen** genoemd.

Laten we hier afspreken dat 'de verkeersonveiligheid' in dit geval wordt gemeten als zijnde het aantal doden en ziekenhuisgewonden zoals door de politie op kruispunten geregistreerd over een bepaalde periode.

2. Doelpopulatie van de maatregel, en steekproeftrekking

De tweede vraag die de bovengenoemde onderzoeksvraag oproept luidt: gaat het hierbij om *alle* soorten van kruispunten? Of alleen kruispunten binnen de bebouwde kom? Of alleen viertaks kruispunten, etc.? Het beantwoorden van dit type vragen komt neer op het bepalen van de **populatie** van ongevallen of slachtoffers waarover

we na afloop van de ex post evaluatie een uitspraak willen kunnen doen. En dit is weer van belang om te bepalen welke **steekproef** van kruispunten in de ex post evaluatie gaat worden gebruikt.

Verder is het – uiteraard – van cruciaal belang om de metingen in het onderzoek te beperken tot doden en ziekenhuisgewonden als gevolg van ongevallen die zich op kruispunten (c.q. rotondes) hebben afgespeeld.

Stel dat we besluiten om de onderzoeksvraag te beperken tot Nederlandse kruispunten buiten de bebouwde kom. Nu precies is vastgesteld wat er onder 'de verkeers(on)veiligheid' wordt verstaan, en om welke populatie van kruispunten het gaat, zouden in een *echt experimenteel* onderzoek typisch de volgende stappen worden doorlopen.

3. Het experimentele design: randomized controlled trial

Eerst wordt er **aselect** een **steekproef** van (bijvoorbeeld) zestig kruispunten getrokken uit de gekozen doelpopulatie van alle Nederlandse kruispunten buiten de bebouwde kom. Het woord aselect houdt in dat ieder kruispunt in de doelpopulatie evenveel kans heeft om in de steekproef terecht te komen. Alleen dan is het toegestaan om de resultaten gevonden voor de steekproef van zestig kruispunten na afloop van het onderzoek te generaliseren naar de populatie bestaande uit alle Nederlandse kruispunten buiten de bebouwde kom.

Vervolgens worden de zestig kruispunten in de steekproef door middel van random (oftewel willekeurige, toevalsgewijze) toewijzing opgedeeld in twee gelijke groepen van dertig kruispunten. De dertig kruispunten in de ene groep worden omgebouwd tot rotondes; deze kruispunten behoren tot de **experimentele groep**. De overige dertig kruispunten worden *niet* omgebouwd tot rotondes; deze kruispunten behoren tot de **controlegroep**.

Het doel en belang van de **random toewijzing** van de kruispunten aan de experimentele en de controlegroep is dat alle mogelijke a priori verschillen tussen de kruispunten die óók samenhangen met hun verkeersveiligheid (etmaalintensiteiten, verlichting, bijvoorbeeld) zo eerlijk mogelijk verdeeld worden over de twee groepen.

Daarna wordt - gedurende een weldoordachte periode - vóór ombouw van de kruispunten in de experimentele groep naar rotondes het aantal overleden en zwaargewonde verkeersslachtoffers op ieder van de zestig onderzochte kruispunten gemeten. Dit wordt de **voormeting** genoemd. Vervolgens worden de dertig kruispunten in de experimentele groep omgebouwd tot rotondes, waarna - weer gedurende een weldoordachte periode - het aantal overleden en zwaargewonde verkeersslachtoffers op ieder van de zestig onderzochte locaties wordt gemeten. Hiervan zijn er nu dertig rotondes (de experimentele groep), en dertig nog steeds kruispunten (de controlegroep). Deze tweede meting wordt de **nameting** genoemd.

Met 'een weldoordachte periode' wordt bedoeld dat de duur van de periode lang genoeg moet zijn om voldoende aantallen slachtoffers te kunnen meten. Verder dienen begin- en einddatum van de periode waarin gemeten wordt uiteraard voor alle locaties gelijk te zijn.

		voormeting	nameting
Experimentele Groep	s1	17	8
	s2	22	2
	:	:	:
	s30	24	12
Controlegroep	s31	8	17
	s32	12	25
	:	:	:
	s60	23	26

Tabel 1. *De ideale proefopzet.*

Na afloop van het experiment kunnen de verkregen gegevens worden weergegeven zoals in Tabel 1. In deze tabel zijn de getallen 17, 22, etcetera, de geobserveerde aantallen overleden en zwaargewonde verkeersslachtoffers. Zij vormen tezamen de zogenaamde **afhankelijke** variabele.

De **onafhankelijke** variabelen in dit experimentele onderzoeksdesign zijn 'tijdstip' (met twee niveaus, namelijk de voor- en de nameting), 'conditie' (met twee niveaus, namelijk de experimentele en de controlegroep), en de interactie tussen 'tijdstip' en 'conditie'.

De letter *s* in de symbolen 's1', 's2', etc. in Tabel 1 staat voor studie object. In dit geval zijn de studie objecten kruispunten.

De gegevens in Tabel 1 kunnen worden geanalyseerd door middel van **variantie-analyse**. Een variantie-analyse wordt ook wel een ANOVA genoemd, de afkorting van het Engelse ANalysis Of VAriance. De resultaten van de variantie-analyse van deze gegevens staan in Tabel 2. In deze tabel worden een aantal statistische toetsen weergegeven waarmee beoordeeld kan worden of allerlei gemiddelden al dan niet significant van elkaar verschillen.

Uit de eerste *F*-toets in Tabel 2 (voor 'tijdstip') blijkt dat het gemiddelde aantal overleden en zwaargewonde verkeersslachtoffers bij de voormeting significant anders is dan bij de nameting, ongeacht of het om kruispunten in de experimentele of in de controlegroep gaat. Deze gemiddelden zijn gelijk aan 19.9 (voormeting) en 14.7 (nameting). Dat deze twee gemiddelden significant van elkaar verschillen is af te lezen aan de waarde van *p* in de laatste kolom van Tabel 2. Deze waarde geeft de kans aan dat de nulhypothese van gelijke gemiddelden ten onrechte wordt verworpen. Voor 'tijdstip' is de waarde van *p* gelijk aan 0.000. Aangezien dit getal kleiner is dan 0.01, is de kans dat de nulhypothese ten onrechte wordt verworpen kleiner dan 1%. We kunnen dus met grote zekerheid concluderen dat de gemiddelden bij voor- en nameting van elkaar verschillen.

Hetzelfde geldt voor de gemiddelden van de experimentele (14.8) en de controlegroep (19.8), ongeacht of het de voor- of de nameting betreft (zie de toets voor 'conditie' in Tabel 2). Ook deze twee gemiddelden zijn significant verschillend op 1% niveau. De toetsen voor 'tijdstip' en 'conditie' zijn dus beide significant. Deze variabelen zijn echter niet van belang voor het beantwoorden van de onderzoeksvraag. In deze onderzoeksopzet is alleen de interactie tussen 'conditie' en 'tijdstip' relevant voor het beantwoorden van de onderzoeksvraag.

Bron	SS	df	MS	F	p
1. tijdstip	826.9	1	826.9	[1/3] = 35.8	0.000
2. conditie x tijdstip	1003.4	1	1003.4	[2/3] = 43.5	0.000
3. error(1)	1339.2	58	23.1		
4. conditie	765.1	1	765.1	[4/5] = 23.1	0.000
5. error(2)	1615.4	58	27.9		
6. totaal	5553.0	119			

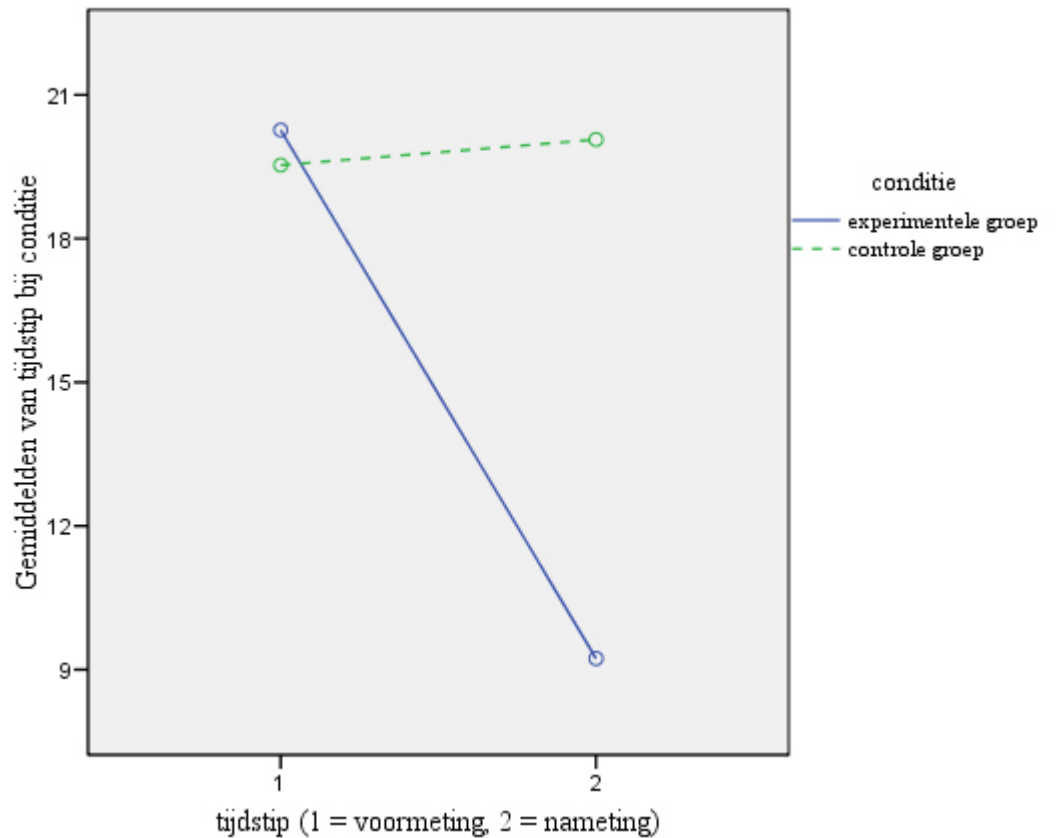
Tabel 2. ANOVA tabel van de gegevens in Tabel 1.

Uit Tabel 2 blijkt dat ook de toets voor de interactie tussen 'conditie' en 'tijdstip' significant is op 1% niveau. Dit betekent dat de verandering tussen voor- en nameting in het aantal overleden en zwaargewonde verkeersslachtoffers in de experimentele groep *anders is verlopen* dan in de controlegroep. Hoe anders is het makkelijkst te zien door deze vier gemiddelden in een grafiek weer te geven, zoals in Figuur 1.

Ten eerste valt op dat de gemiddelden van de experimentele en de controlegroep bij de voormeting vrijwel gelijk zijn aan elkaar. Dat is het resultaat van de random toewijzing van de zestig kruispunten aan de experimentele en controlegroep. Verder is in Figuur 1 duidelijk te zien dat de gemiddelden van de controlegroep bij de voor- en nameting vrijwel gelijk zijn gebleven, terwijl het gemiddelde van de experimentele groep bij de nameting lager is dan bij de voormeting.

Hieruit blijkt dat het ombouwen van de dertig kruispunten naar rotondes gepaard is gegaan met een daling van het aantal overleden en zwaargewonde verkeersslachtoffers.

Een zeer prettige eigenschap van dit ideale experimentele design is dat bovendien ook geconcludeerd mag worden dat de maatregel (het ombouwen van kruispunten naar rotondes) de gevonden daling in het aantal overleden en zwaargewonde verkeersslachtoffers heeft *veroorzaakt*. In de experimentele conditie is immers aangetoond dat de maatregel tot verandering heeft geleid; in de controle conditie is bewezen dat er zonder maatregel geen verandering optreedt. Gecombineerd is hiermee hard gemaakt dat de maatregel een voldoende en noodzakelijke voorwaarde is voor het doen afnemen van het aantal verkeersslachtoffers. Vormen de zestig kruispunten een aselechte steekproef uit de populatie van Nederlandse kruispunten buiten de bebouwde kom, dan is het tenslotte ook toegestaan om de conclusies uit dit onderzoek te generaliseren naar alle Nederlandse kruispunten buiten de bebouwde kom. De eindconclusie bij deze (fictieve) gegevens luidt dan ook dat het ombouwen van kruispunten buiten de bebouwde kom naar rotondes resulteert in een daling van het aantal overleden en zwaargewonde verkeersslachtoffers.



Figuur 1. Gemiddelden van interactie tussen tijdstip en conditie.

Is de interactie tussen conditie en tijdstip niet significant, dan moet geconcludeerd worden dat er in het onderzoek geen effect van de invoering van de betreffende maatregel is gevonden. Dit kan weer twee oorzaken hebben:

1. de maatregel sorteert daadwerkelijk geen effect;
2. de maatregel sorteert wel degelijk effect maar de variatie in de observaties is dusdanig groot dat het effect van de maatregel erin ondergesneeuwd is geraakt. Het laatste punt hangt samen met de **power** van een toets. Dat is de kans dat een toets een verschil detecteert als dat verschil er wel degelijk is. In het algemeen kan de power van een toets worden verhoogd door de steekproef te vergroten (betrek meer kruispunten in het onderzoek), dan wel door de variabiliteit in de observaties te verkleinen middels experimentele controle (beperk het onderzoek tot alleen die kruispunten met een bepaalde range aan verkeersintensiteiten bijvoorbeeld).

4. Alternatieve experimentele designs en gevolgen voor de conclusies

We zullen nu een aantal alternatieve onderzoeksopzetten bespreken die ook gebruikt kunnen worden (en gebruikt worden) voor het beantwoorden van de eerder genoemde onderzoeksvraag. We bespreken daarbij steeds wat het verschil is met de ideale opzet, en wat de gevolgen zijn voor de conclusies die uit de resultaten van het onderzoek kunnen worden getrokken.

4.1 Experiment zonder voormeting

De eerste alternatieve proefopzet is identiek aan die welke besproken werd in paragraaf 3 met als enige verschil dat er geen voormeting wordt uitgevoerd (Tabel 3).

		Nameting
Experimentele Groep	s1	8
	s2	2
	:	:
	S30	12
Controlegroep	s31	17
	s32	25
	:	:
	S60	26

Tabel 3. *Geen voormeting.*

In deze onderzoekopzet is er slechts één onafhankelijke variabele: 'conditie'. Bij de nameting is het gemiddelde aantal overleden en zwaargewonde verkeersslachtoffers in de experimentele groep 9.2, en in de controlegroep 20.1.

Bron	SS	df	MS	F	p
1. conditie	1760.4	1	1760.4	[1/2] = 76.4	0.000
2. error	1337.2	58	23.1		
3. totaal	3097.6	59			

Tabel 4. *ANOVA tabel van de gegevens in Tabel 3.*

Zoals blijkt uit Tabel 4 is de toets voor 'conditie' significant op 1% niveau ($p < 0.01$), zodat de nulhypothese van gelijke gemiddelden in experimentele en controlegroep kan worden verworpen.

Mag nu geconcludeerd worden dat de maatregel heeft geleid tot een daling in het aantal slachtoffers?

Het antwoord is: ja, maar alleen als de zestig kruispunten random zijn toegewezen aan de experimentele en controlegroep. Indien er geen random toewijzing aan de twee condities heeft plaatsgevonden dan is een alternatieve verklaring voor het gevonden effect dat de experimentele groep en de controlegroep op andere kenmerken verschillen die ook samenhangen met de afhankelijke variabele. Stel bijvoorbeeld dat de etmaalintensiteiten op de kruispunten in de controlegroep toevallig veel hoger zijn dan op de rotondes in de experimentele groep, dan vormen deze verschillen een mogelijke alternatieve verklaring voor het gevonden effect. Zulke kenmerken worden **confounding** of **verstorende** variabelen genoemd. Voor dergelijke verschillen kan overigens in de analyse statistisch worden gecontroleerd met behulp van co-variantie analyse, maar *alleen als deze verstorende variabelen bekend én gemeten zijn*.

We merken verder op dat - mocht er geen random toewijzing van de studie objecten hebben plaatsgevonden aan de experimentele en controle condities - het prettige van een voormeting is dat we dan tenminste nog kunnen corrigeren voor eventuele

aanvangsverschillen tussen de studie objecten door de voormeting als covariaat mee te nemen in de variantie-analyse (zie bijvoorbeeld Twisk, Vlakveld en Commandeur, 2007).

4.2. Experiment zonder controlegroep

Een andere alternatieve onderzoeksopzet is weergegeven in Tabel 5. In dit onderzoeksdesign is er wel een voor- en een nameting, maar geen controlegroep.

		voormeting	nameting
Experimentele Groep	s1	17	8
	s2	22	2
	:	:	:
	s30	24	12

Tabel 5. *Geen controlegroep.*

De enige onafhankelijke variabele is nu dus 'tijdstip'. Het gemiddelde aantal slachtoffers tijdens de voormeting is 20.3, dat tijdens de nameting is 9.2. Het resultaat van de variantie-analyse van deze gegevens wordt in Tabel 6 getoond. De toets voor 'tijdstip' is significant op 1% niveau ($p < 0.01$), zodat geconcludeerd kan worden dat het ombouwen van de dertig kruispunten in rotondes gepaard is gegaan met een daling in het aantal verkeersslachtoffers.

<i>Bron</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>
1. tijdstip	1826.0	1	1826.0	[1/2] = 94.6	0.000
2. error(1)	559.5	29	19.3		
3. error(2)	687.8	29			
4. totaal	3073.3	59			

Tabel 6. *ANOVA tabel van de gegevens in Tabel 5.*

Vraag is weer: mag hieruit worden geconcludeerd dat de maatregel een daling in het aantal verkeersslachtoffers heeft *veroorzaakt*? Het antwoord is: nee, want de gevonden daling kan net zo goed door iets anders dan de maatregel zijn veroorzaakt, en wel door

1. een algemeen dalende trend;
 2. regressie naar het gemiddelde;
 3. verandering in registratie;
 4. andere externe gebeurtenissen dan de maatregel;
- of een combinatie van deze.

De eerste alternatieve verklaring voor een gevonden effect hangt samen met het feit dat het aantal overleden en zwaargewonde verkeersslachtoffers in Nederland sinds begin jaren zeventig van de vorige eeuw gestaag aan het afnemen is. In het laatstgenoemde design kan dit een (gedeeltelijke) verklaring vormen voor het gevonden verschil tussen voor- en nameting.

De tweede verklaring komt om de hoek kijken als de dertig onderzochte kruispunten bij de voormeting al een bovengemiddeld aantal verkeersslachtoffers laten zien. In dat geval kan het aantal slachtoffers bij de tweede meting alleen al op grond van toeval richting het gemiddelde zijn opgeschoven, hetgeen op zich al leidt tot een

schijnbare afname in het aantal slachtoffers. Dit verschijnsel wordt voluit **statistische regressie naar het gemiddelde** genoemd.

Bij de derde verklaring zou het aantal slachtoffers alleen schijnbaar kunnen zijn afgenomen omdat de politie bij de nameting minder accuraat is gaan registreren dan bij de voormeting.

Als voorbeeld voor de vierde alternatieve verklaring voor een geobserveerde afname in verkeersslachtoffers zou kunnen dienen dat de benzineprijzen tussen voor- en nameting flink zijn gestegen, waardoor mensen zich minder vaak per auto zijn gaan verplaatsen. Ook bij een experiment zonder controlegroep geldt trouwens dat er voor zulke confounding of verstoringe variabelen statistisch gecorrigeerd kan worden middels co-variantie analyse, mits deze verstoringe variabelen bij voor- en nameting bekend én gemeten zijn. Zelfs dan blijft in een experiment zonder controlegroep altijd onduidelijk of er geen andere onvermoede externe gebeurtenissen mogelijk aan een geconstateerde verandering hebben bijgedragen.

Het algemene probleem van een simpele voor- en nastudie is, kortom: we hebben geen idee wat de nasituatie zou zijn geweest als de maatregel niet was ingevoerd. De aanname dat er zonder maatregel tussen voor- en nameting niets zou zijn veranderd is op zijn zachtst gezegd aanvechtbaar.

Dit toont het grote belang aan van het opnemen van een controlegroep in het onderzoeksdesign, én van het random toewijzen van de kruispunten aan de twee condities. Als er sprake is van een algemeen dalende trend, van statistische regressie, van een verandering in registratie, van andere externe gebeurtenissen, dan wel van een combinatie van deze vier, dan zullen door de random toewijzing de kruispunten in de experimentele groep en de kruispunten in de controlegroep tussen voor- en nameting *in gelijke mate beïnvloed worden door deze factoren*. Is de interactie tussen 'conditie' en 'tijdstip' in dat geval significant, dan resteert er nog maar één mogelijke verklaring voor het gevonden effect, en dat is de maatregel zelf.

Het zal tenslotte duidelijk zijn dat een simpele voor- en nastudie waarbij de verandering in aantallen slachtoffers voor en na ombouw van slechts één kruispunt naar een rotonde wordt onderzocht wel erg veel ruimte laat voor kritiek op de conclusie dat de maatregel effect heeft gesorteerd. Als in zo'n onderzoek een daling in het aantal slachtoffers wordt gevonden dan is de alternatieve verklaring voor het effect van de maatregel: de daling is puur toevallig. Voor meer informatie over experimentele proefopzetten en (co)variantie-analyse kan bijvoorbeeld Kirk (1968) worden geraadpleegd.

Samenvattend vereist de ideale onderzoeksopzet voor de ex post evaluatie van de effecten van maatregelen de volgende experimentele manipulaties:

1. een random trekking van de steekproef van studie objecten uit de doelpopulatie (de populatie waarop de maatregel geacht wordt effect te sorteren);
2. een random toewijzing van de studie objecten in de steekproef aan twee experimentele condities: een groep studie objecten die de maatregel wel ondergaan, en een groep studie objecten die de maatregel niet ondergaan.

De eerste vereiste maakt dat de conclusies uit het onderzoek (welke deze ook zijn) gegeneraliseerd mogen worden naar de totale doelpopulatie omdat de random trekking garandeert dat de steekproef representatief is voor de doelpopulatie. De

tweede vereiste maakt het mogelijk om te concluderen dat de maatregel een effect (indien gevonden) heeft *veroorzaakt* omdat de random toewijzing aan de twee condities garandeert dat mogelijke versturende variabelen gelijkelijk over de twee condities zijn verdeeld.

5. Observatie onderzoek

5.1 Inleiding

Meestal is het in verkeersveiligheidsonderzoek om praktische of ethische redenen niet mogelijk om onderzoeksubjecten random te selecteren uit de doelpopulatie, dan wel om onderzoeksubjecten random toe te wijzen aan een experimentele en een controlegroep. Een voorbeeld is de onderzoeksvraag of autobestuurders onder invloed van alcohol, drugs of medicijnen (of een combinatie van deze) een verhoogd risico hebben om bij een ongeval betrokken te raken. Het zal duidelijk zijn dat het in dit geval niet mogelijk is om een aselechte steekproef van autobestuurders random toe te wijzen aan de experimentele en de controlegroep, en de bestuurders in de experimentele groep vervolgens onder invloed van alcohol, drugs of medicijnen de weg op te sturen. Meer in het algemeen is het om ethische redenen niet mogelijk om proefpersonen random toe te wijzen aan omstandigheden die kunnen resulteren in menselijk lijden.

Veel onderzoeken naar de effecten van verkeersveiligheidsmaatregelen zijn om die redenen dan ook niet gebaseerd op experimentele onderzoeksdesigns maar puur observationeel van aard. In tegenstelling tot de in paragraaf 3 beschreven ideale experimentele opzet kunnen er echter geen harde oorzaak-gevolg conclusies aan dergelijke observationele studies worden verbonden.

Desondanks is het de kunst om ook in deze situatie de ideale onderzoeksopzet zoveel mogelijk te trachten te benaderen door:

- te proberen vast te stellen wat de verkeersveiligheidssituatie zou zijn geweest indien de maatregel niet was ingevoerd;
- statistisch in plaats van experimenteel te controleren voor versturende variabelen; de (steekproef uit de) doelpopulatie te vergelijken met een (steekproef uit een) referentiepopulatie waarop de onderzochte maatregel geacht wordt geen effect te hebben gesorteerd, maar die verder zoveel mogelijk dezelfde kenmerken heeft als de doelpopulatie.

In paragraaf 6 laten we zien hoe dit alles kan worden bewerkstelligd met tijdreeks analyse, en in paragraaf 7 met de zogenoemde Empirical Bayes methode.

6. Tijdreeks analyse

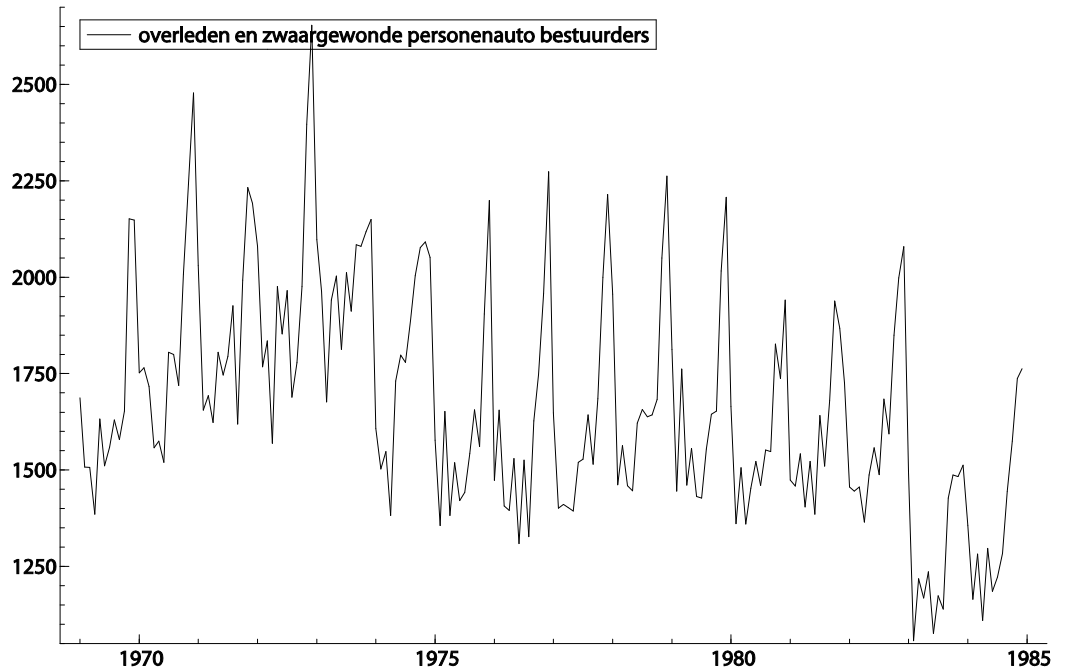
6.1 Inleiding

We lichten een aantal belangrijke methodologische aspecten van het ex post evalueren van verkeersveiligheidsmaatregelen toe aan de hand van de volgende observationele studie (zie Harvey en Durbin, 1986). In Groot-Brittannië werden bestuurders van personenauto's en passagiers voorin de auto vanaf 31 januari 1983 wettelijk verplicht om een autogordel te dragen. Het percentage gordel dragers bedroeg 40% in december 1982; in februari 1983 was dit percentage gestegen naar 90% en stabiliseerde vanaf maart 1983 op ongeveer 95%.

Het dragen van de gordel wordt verondersteld alleen effect te hebben op het aantal slachtoffers van personenauto ongevallen en hun letselernst, niet op het aantal personenauto ongevallen zelf. Mogelijke versturende variabelen die ook van invloed

zijn op de letselernst van slachtoffers van auto ongevallen zijn: snelheid, type auto, soort ongeval, en het aantal inzittenden (Elvik en Vaa, 2004).

Beperken we ons voorlopig even tot de bestuurders van personenauto's dan zijn de maandgegevens van de aantallen doden en ernstig gewonden voor de jaren 1969 tot en met 1984 weergegeven in Figuur 2.



Figuur 2. Maandcijfers van in Groot-Brittannië overleden en zwaargewonde personenauto bestuurders in de periode 1969-1984.

De simpelste maar ook meest naïeve evaluatie van het effect van de gordeldracht zou nu zijn om bijvoorbeeld de totale maandgegevens van februari 1982 tot en met januari 1983 (de 'voormeting') te vergelijken met de totale maandgegevens van februari 1983 tot en met januari 1984 (de 'nameting'), zie Tabel 7.

1982(2)-1983(1)	1983(2)-1984(1)	Totaal
19498	15335	34833

Tabel 7. Totaal aantal overleden en zwaargewonden autobestuurders in het jaar voor en na invoering van de gordeldracht in februari 1983 in Groot-Brittannië.

De F-toets voor het verschil tussen deze twee frequenties is

$$F = \frac{N_1}{N_2 + 1} = \frac{19498}{15335 + 1} = 1.27,$$

met $2(N_2 + 1) = 30672$ vrijheidsgraden in de teller, en $2N_1 = 38996$ vrijheidsgraden in de noemer (zie Kanji, 1993, p.51). De toets is zeer significant ($p < 0.01$),

hetgeen betekent dat de nulhypothese van gelijke frequenties kan worden verworpen. We concluderen dat de invoering van de gordeldraagplicht geassocieerd is met een $100((19498-15335)/19498) = 21.35\%$ afname in het aantal overleden en zwaargewonde autobestuurders.

Ook kunnen we alle maandcijfers (dat zijn er 169) in Figuur 2 van vóór de invoering van de gordelplicht in februari 1983 bij elkaar optellen, alle maandcijfers (dat zijn er 23) van na de invoering van de gordelplicht bij elkaar optellen, en het verschil tussen deze twee optellingen toetsen, gecorrigeerd voor de verschillen in tijdsduur waarover de optellingen zijn gemaakt:

$$F = \frac{\frac{1}{t_1}(N_1 + 0.5)}{\frac{1}{t_2}(N_2 + 0.5)} = \frac{\frac{1}{169}(290308 + 0.5)}{\frac{1}{23}(30399 + 0.5)} = 1.30,$$

met $(2N_1+1)$ vrijheidsgraden in de teller, en $(2N_2+1)$ vrijheidsgraden in de noemer (Kanji, 1993, p.51). Ook deze toets is zeer significant ($p < 0.01$), en wijst op een afname van 23%.

Vraag is nu: is dit een overtuigend resultaat? Alle bezwaren die we in paragraaf 4.2 hebben genoemd zijn in deze situatie van kracht: de gevonden daling kan net zo goed zijn veroorzaakt door een algemeen dalende trend, door regressie naar het gemiddelde, door veranderingen in registratie, door andere externe gebeurtenissen dan de invoering van de gordelplicht, of door een combinatie van deze.

Door tijdreeks analyses van de gegevens in Figuur 2 met behulp van state space modellen (zie Harvey, 1989, Durbin en Koopman, 2001, Commandeur en Koopman, 2007) is het onder andere mogelijk om een inschatting te krijgen van wat de ontwikkeling zou zijn geweest als de gordelplicht niet was ingevoerd. Dit wordt in paragraaf 6.2 besproken. Daarnaast kunnen we met een tijdreeks analyse het effect van de maatregel bepalen gecorrigeerd voor een mogelijke trend in de gegevens, voor maandschommelingen, en voor verstorende variabelen, zie paragraaf 6.3. Tenslotte wordt in paragraaf 6.4 getoond hoe gebruik kan worden gemaakt van iets dat lijkt op de controle groep in een randomized controlled trial experiment om impliciet te corrigeren voor mogelijk verstorende variabelen.

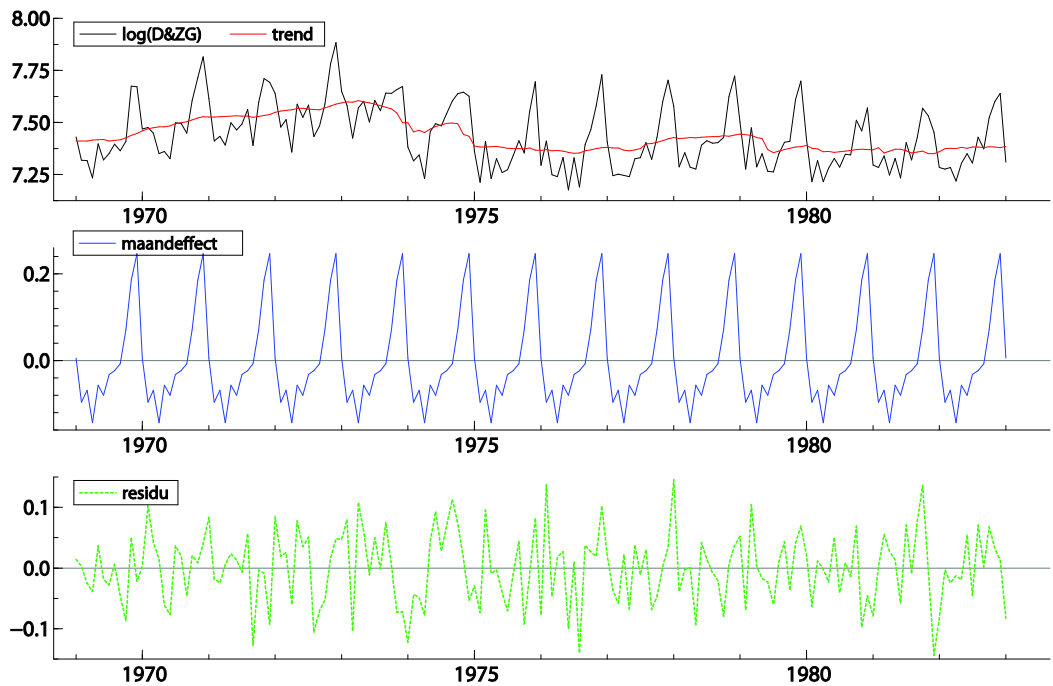
6.2. Prognoses van ontwikkelingen zonder maatregel

Om een inschatting te krijgen van wat de ontwikkeling zou zijn geweest als de gordelplicht niet was ingevoerd analyseren we alleen de gegevens in Figuur 2 van januari 1969 tot en met januari 1983 vóór invoering van de gordelplicht (dit zijn 169 maanden), en wel met het volgende structurele tijdreeksmodel

$$\log(y_t) = \mu_t + \gamma_t + \varepsilon_t \quad (1)$$

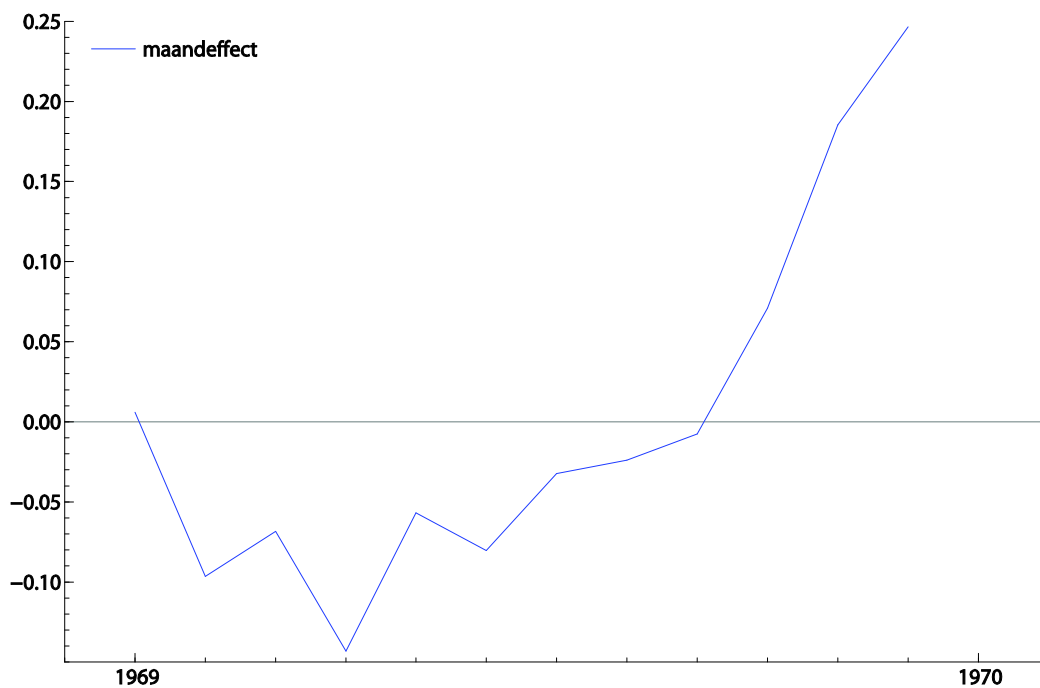
voor $t = 1, \dots, 169$. In deze vergelijking is y_t het aantal doden en zwaargewonden in maand t , modelleert μ_t de (mogelijk tijdvariërende) trend in de tijdreeks, modelleert γ_t het (mogelijk tijdvariërende) maandeffect in de reeks, en is ε_t het residu van het model (d.w.z. dat deel van een observatie dat niet door het model wordt verklaard). Voor de gedetailleerde technische aspecten van de state space

methodologie verwijzen we naar de eerder genoemde literatuur. Dit model blijkt tot een goede weergave van de geobserveerde ontwikkelingen tot en met januari 1983 te leiden; het resultaat van deze analyse is te zien in Figuur 3.



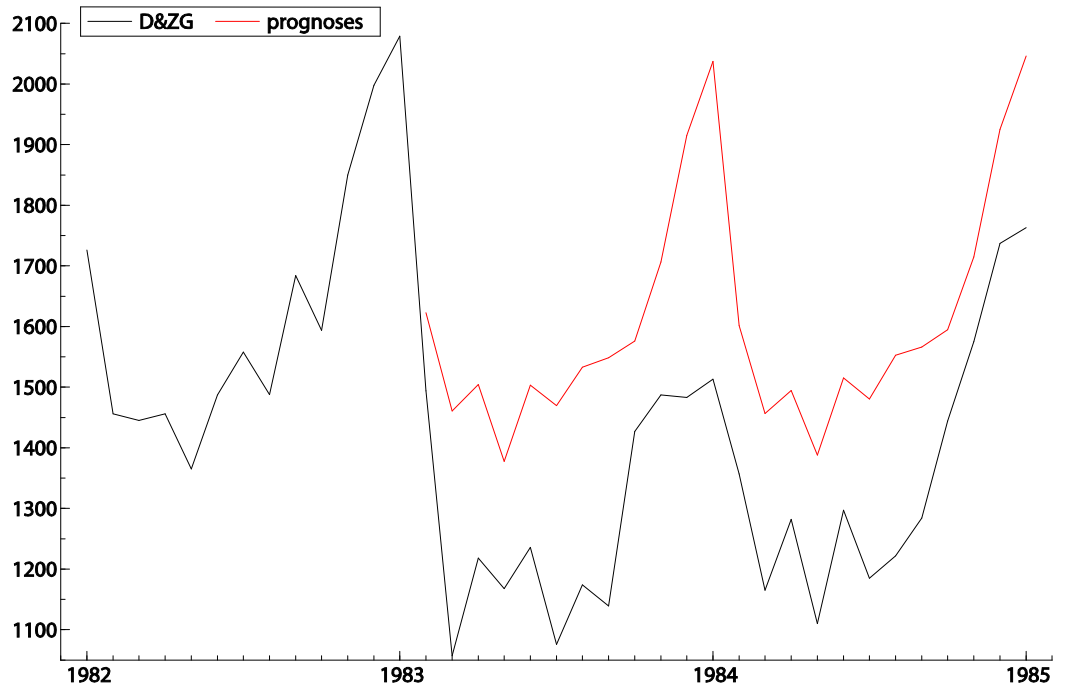
Figuur 3. Resultaat tijdreeks analyse van de logaritme van het aantal Britse doden en zwaargewonden (D&ZG) in de periode 1969 tot en met januari 1983.

In Figuur 4 worden de maandschommelingen in Figuur 3 apart getoond voor het jaar 1969. Aangezien het maandeffect bij deze tijdreeks voor alle jaren identiek blijkt te zijn, kunnen we concluderen dat de maanden november en nog meer december bij autobestuurders in Groot-Brittannië gepaard gaan met de hoogste aantallen verkeersslachtoffers, terwijl de maand april voor deze groep weggebruikers steeds tot de minste slachtoffers leidt.



Figuur 4. *Het maandeffect voor het aantal Britse doden en zwaargewonden (D&ZG) in 1969.*

Op grond van dit model kunnen we vervolgens prognoses berekenen voor de periode februari 1983 tot en met december 1984, en deze prognoses vergelijken met de waargenomen aantallen slachtoffers. Het resultaat is te zien in Figuur 5.



Figuur 5. De vanaf februari 1983 geprognosticeerde aantallen Britse doden en zwaargewonden (rode lijn), alleen gebaseerd op de geobserveerde ontwikkelingen van januari 1969 tot en met januari 1983. De werkelijke aantallen zijn met de zwarte lijn weergegeven.

Duidelijk is dat het waargenomen aantal (de zwarte lijn) doden en zwaargewonden bij autobestuurders in de periode februari 1983 tot en met december 1984 kleiner is dan het aantal (de rode lijn) dat we op grond van het verleden tot en met januari 1983 zouden mogen verwachten.

prognoses	observaties
36970	15335

Tabel 8. Geprognosticeerd en geobserveerd totaal aantal overleden en zwaargewonden autobestuurders in de periode februari 1983-december 1984 in Groot-Brittannië.

De totale aantallen voor de periode februari 1983 tot en met december 1984 zijn in Tabel 8 weergegeven. De *F*-toets voor het verschil tussen de frequenties in Tabel 8 is

$$F = \frac{N_1}{N_2 + 1} = \frac{36970}{15335 + 1} = 2.41,$$

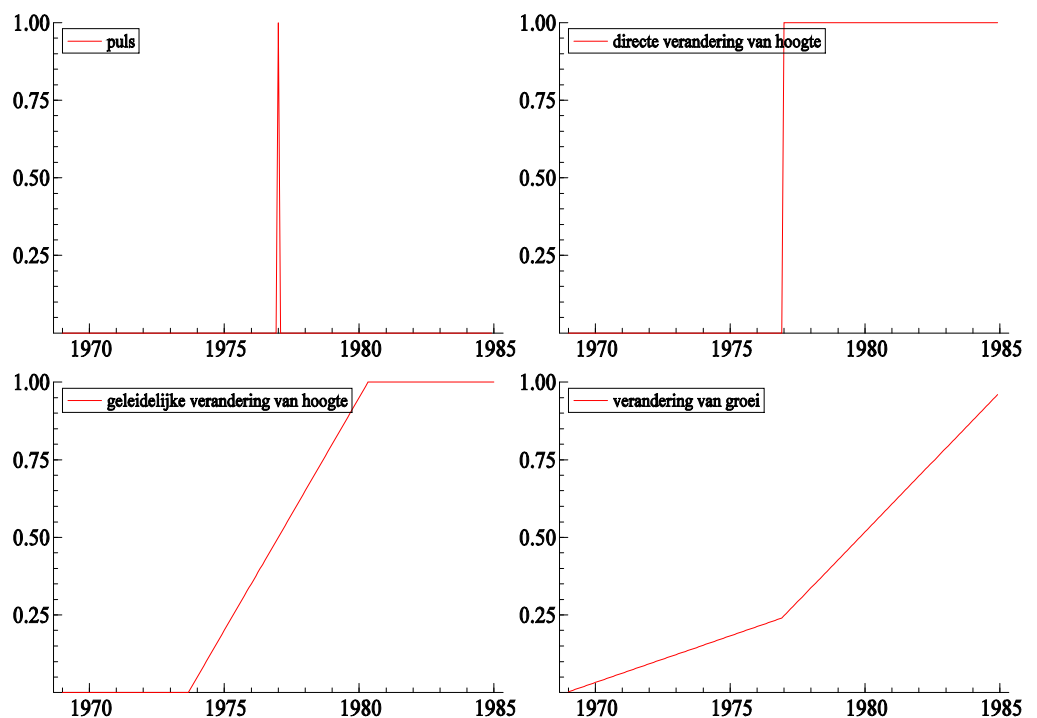
met $2(N_2+1) = 30672$ vrijheidsgraden in de teller, en $2N_1 = 73940$ vrijheidsgraden in de noemer, een zeer significant verschil ($p < 0.01$). Dit is een eerste aanwijzing dat er zich in februari 1983 een belangrijke verandering in de ontwikkeling heeft voorgedaan.

6.3 Interventie analyse gecorrigeerd voor verstorende variabelen

Vervolgens analyseren we de complete tijdreeks, en voegen een **interventie variabele** toe aan model (1):

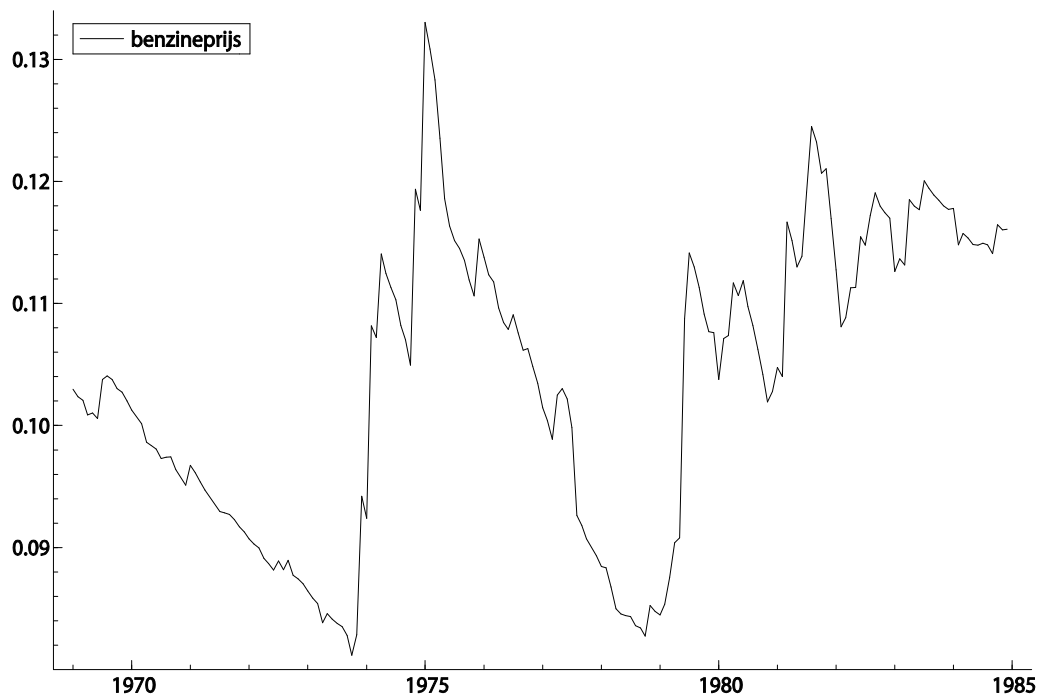
$$\log(y_t) = \mu_t + \gamma_t + \lambda\omega_t + \varepsilon_t \quad (2)$$

voor $t = 1, \dots, 192$. Hierin is λ een onbekend (te schatten) regressiegewicht, en ω_t de interventie variabele. Maatregelen kunnen in principe verschillende effecten sorteren, een aantal waarvan in Figuur 6 worden getoond. Een maatregel kan ervoor zorgen dat het aantal ongevallen of slachtoffers zeer tijdelijk afneemt waarna de tijdreeks onmiddellijk weer terugkeert naar het oude niveau. Dit wordt een puls genoemd. Een maatregel kan leiden tot een directe en blijvende afname (rechtsboven in de figuur) dan wel geleidelijke en blijvende afname (linksonder in de figuur) in het aantal ongevallen of slachtoffers. Bovendien kan een maatregel ervoor zorgen dat de *groei* of *daling* in het aantal ongevallen of slachtoffers toe- dan wel afneemt (rechtsonder in de figuur). Tenslotte is het ook mogelijk om een interventie te modelleren waarbij het aantal ongevallen of slachtoffers direct afneemt, maar na verloop van tijd geleidelijk aan weer terugkeert naar het oude niveau (niet getoond in de figuur).



Figuur 6. Soorten interventie.

In het geval van evaluatie van de gordelplicht kiezen we voor een directe en blijvende verandering van hoogte door de variabele ω_t in (2) te coderen met nullen tijdens de eerste 169 maanden (vóór februari 1983) en met enen in de laatste 23 maanden (tijdens en na februari 1983).



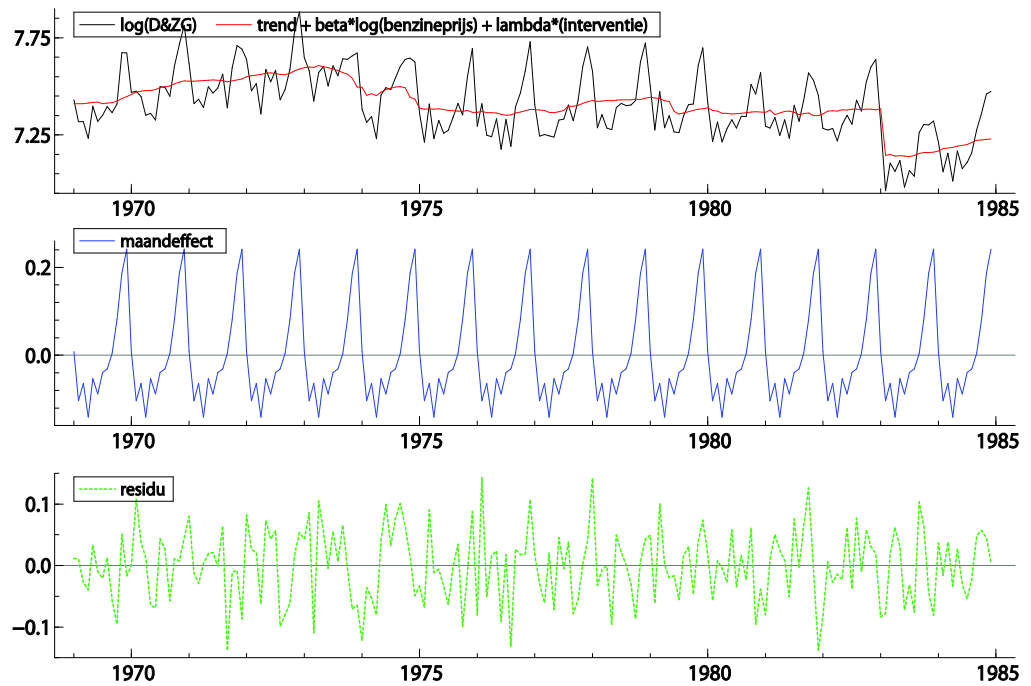
Figuur 7. Britse maandcijfers voor benzineprijs in de periode 1969-1984.

Verder voegen we ook nog de logaritme van de in Figuur 7 getoonde tijdreeks van maandelijkse benzineprijzen toe aan model (2):

$$\log(y_t) = \mu_t + \gamma_t + \lambda\omega_t + \beta x_t + \varepsilon_t \quad (3)$$

voor $t = 1, \dots, 192$. Hierin is β weer een onbekend regressiegewicht, en x_t de tijdreeks met de logaritme van de benzineprijzen. Het idee is dat we de benzineprijs kunnen opvatten als een proxy voor de mobiliteit, waarna het effect van de gordelplicht geëvalueerd kan worden na correctie voor veranderingen in mobiliteit.

Het grafische resultaat van deze analyse is weergegeven in Figuur 8.

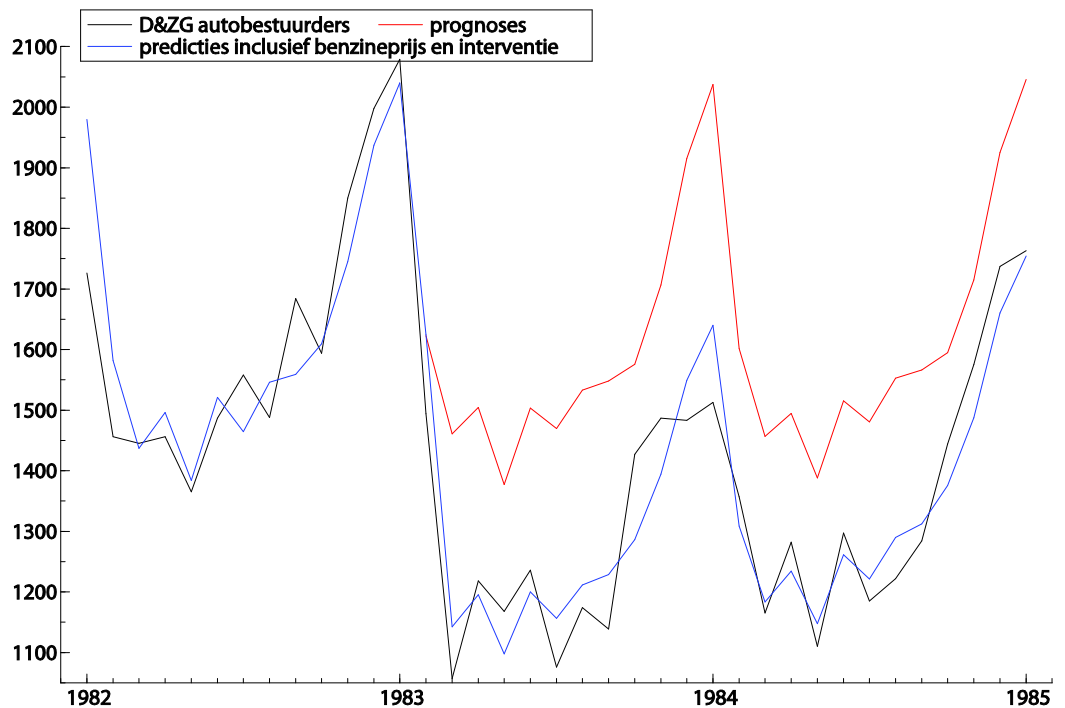


Figuur 8. Resultaat tijdreeks analyse van de logaritme van het aantal Britse doden en zwaargewonden (D&ZG) in de periode januari 1969 tot en met december 1984, inclusief interventie variabele voor de invoering van de gordelplicht en de logaritme van de benzineprijs.

Het regressiegewicht voor de logaritme van de benzineprijs is $\beta = -0.2767$, en wijkt significant af van nul (zie Tabel 9). Dit betekent dat een toename van 1% in benzineprijs in de periode 1969-1984 samengaat met een 0.28% afname in het aantal doden en zwaargewonden bij Britse autobestuurders. Het regressiegewicht voor de interventie variabele is $\lambda = -0.2376$, en wijkt eveneens significant af van nul. Dit betekent dat de introductie van de gordelplicht in februari 1983 – gecorrigeerd voor de ontwikkelingen in benzineprijs - geassocieerd is met een $100(\exp(-0.2376) - 1) = -21.1\%$ verandering in het aantal doden en zwaargewonden bij Britse autobestuurders.

	regressiegewicht	standaardfout	t-waarde	p-waarde
β	-0.2767	0.0984	-2.8122	0.0054
λ	-0.2376	0.0465	-5.1154	7.7e-007

Tabel 9. Effecten van benzineprijs en invoering van de gordelplicht op het aantal overleden en zwaargewonde Britse autobestuurders.



Figuur 9. De vanaf februari 1983 geprognosticeerde aantallen Britse doden en zwaargewonden (rode lijn) alleen gebaseerd op de geobserveerde ontwikkelingen van januari 1969 tot en met januari 1983, samen met de gemodelleerde ontwikkelingen van de volledige tijdreeks inclusief het voor benzineprijs gecorrigeerde effect van de invoering van de gordelplicht (blauwe lijn). De werkelijke aantallen zijn met de zwarte lijn weergegeven.

Het voorafgaande is nog eens grafisch samengevat in Figuur 9. Duidelijk is niet alleen te zien hoe het modelmatig doortrekken van de situatie in de periode tot februari 1983 tot een heel andere ontwikkeling leidt dan wat er daarna daadwerkelijk gebeurde, maar ook hoe het toevoegen van een interventie variabele aan de volledige tijdreeks juist resulteert in een goede representatie van de na januari 1983 geconstateerde ontwikkelingen. Al deze constatering verhogt de aannemelijkheid van de conclusie dat de invoering van de gordelplicht inderdaad effect heeft gehad op de aantallen slachtoffers onder autobestuurders, en gepaard is gegaan met een reductiepercentage van ongeveer 21%.

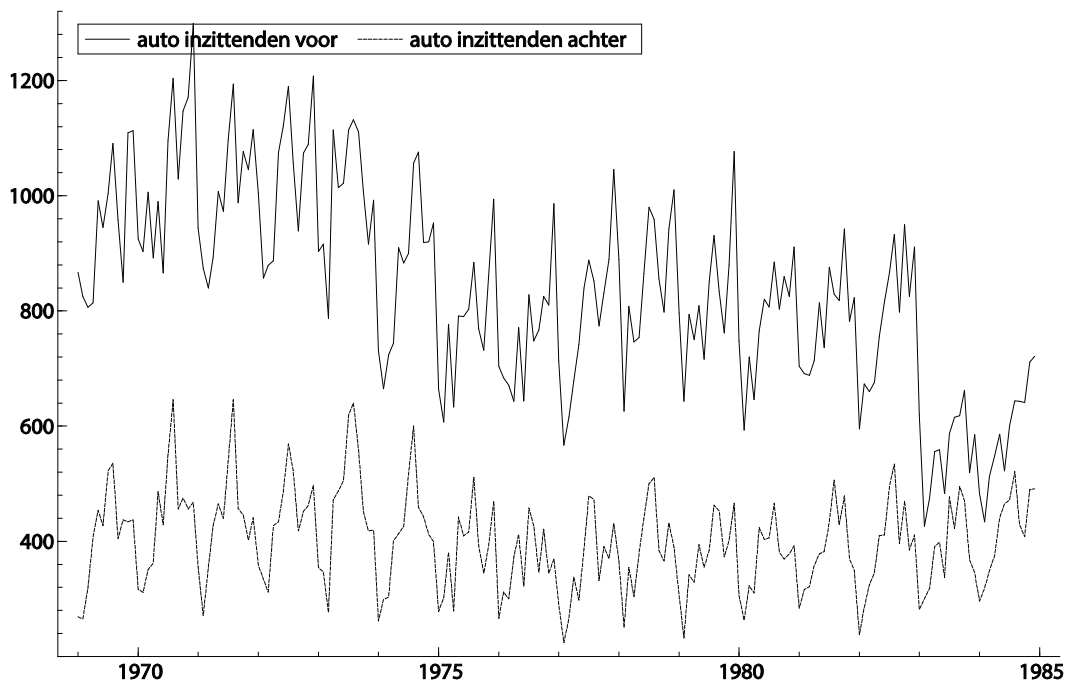
Hiermee is zeker niet bewezen dat de gordelplicht deze reductie heeft veroorzaakt. Maar we hebben tenminste geprobeerd vast te stellen wat er zou zijn gebeurd als de gordelplicht niet was ingevoerd, en we hebben het effect van de maatregel tenminste vastgesteld na correctie voor maandeffecten, voor schommelingen in de benzine prijs en voor een mogelijke algemeen dalende trend in de ontwikkeling van de verkeersveiligheid.

Door gebruik te maken van een lange reeks van 169 observaties vóór de invoering van de maatregel hebben we tenslotte ook aannemelijk gemaakt dat de gevonden reductie niet het gevolg kan zijn van een toevallig eenmalig extreem hoog (of laag!) aantal ongevallen of slachtoffers bij de voormeting, maar dat dit extreem hoog (of laag) aantal ongevallen of slachtoffers de werkelijke en structurele situatie van de

onderzochte groep van weggebruikers weerspiegelt. Hiermee wordt de kans op een regressie naar het gemiddelde effect bij het evalueren van de maatregel wel erg onwaarschijnlijk, zo niet geheel voorkomen.

6.4 Interventie analyse met een referentiegroep

Nog overtuigender wordt het effect van een maatregel als we kunnen aantonen dat de maatregel wel effect heeft op de doelpopulatie, maar geen effect sorteert op een alternatieve populatie die zoveel mogelijk lijkt op de doelpopulatie behalve dan dat de maatregel er niet van op toepassing is. Bestaat er een dergelijke alternatieve populatie, en beschikken we over hun slachtoffergegevens dan kunnen we hiermee de ideale onderzoekopzet benaderen door deze groep als controle groep te gebruiken. Om verwarring met een strikt door random toewijzing verkregen controle groep te voorkomen, zullen we een dergelijke controle groep – in navolging van Hauer (1997) - bij observatie onderzoek een *referentiegroep* noemen.



Figuur 10. Britse maandgegevens van de aantallen overleden en zwaargewonde passagiers voor (boven) en achterin (onder) de auto voor de periode januari 1969 tot en met december 1984.

De invoering van de gordelplicht in februari 1983 in Groot-Brittannië betraf alleen de autobestuurders en passagiers voorin de auto, niet de passagiers achterin de auto. Niet alleen de aantallen overleden en zwaargewonde autobestuurders zijn bekend maar ook de aantallen overleden en zwaargewonde passagiers voorin de auto en achterin de auto. De twee laatstgenoemde tijdreeksen zijn weergegeven in Figuur 10.

Met deze gegevens wordt het mogelijk om dichterbij te komen bij iets dat begint te lijken op een experimenteel design. Om te beginnen kunnen we bijvoorbeeld kijken

naar de frequenties bij voor- en nameting in de 'behandelde' groep en in de referentiegroep, zie Tabel 10.

	1982(2)- 1983(1)	1983(2)- 1984(1)	totaal
inzittenden voor	9482	6568	16050
inzittenden achter	4749	4618	9367
totaal	14231	11186	25417

Tabel 10. Aantallen overleden en zwaargewonde passagiers voor- en achterin de auto in het jaar voor en na invoering van de gordelplicht.

De Chi-kwadraat van deze kruistabel is 168.5 hetgeen wijst op een zeer significante interactie tussen tijdstip en conditie ($p < 0.01$). Als we aannemen dat de verandering in de eerste reeks zonder interventie gelijk zou zijn aan de verandering in de tweede reeks zouden we bij de nameting 9482 (4618/4749) = 9220 overleden en zwaargewonde inzittenden voorin de auto verwachten. Dit suggereert een 100 [(6568-9220)/6568] = -40% reductie als gevolg van de invoering van de gordelplicht. Het bezwaar van deze aanpak is dat we niet corrigeren voor mogelijke trends in de gegevens, noch kunnen corrigeren voor versturende variabelen.

Dit is wel mogelijk met tijdreeks analyse. Een eerste bivariate analyse van de twee tijdreeksen in Figuur 10 met het volgende model

$$\begin{aligned} \log(y_t^{voor}) &= \mu_t^{voor} + \gamma_t^{voor} + \lambda^{voor} \omega_t + \beta^{voor} x_t + \varepsilon_t^{voor} \\ \log(y_t^{achter}) &= \mu_t^{achter} + \gamma_t^{achter} + \lambda^{achter} \omega_t + \beta^{achter} x_t + \varepsilon_t^{achter} \end{aligned} \quad (4)$$

voor $t = 1, \dots, 192$ - waarbij aan beide reeksen de verklarende variabelen invoering gordelplicht en log(benzineprijs) worden toegevoegd - levert de volgende resultaten op.

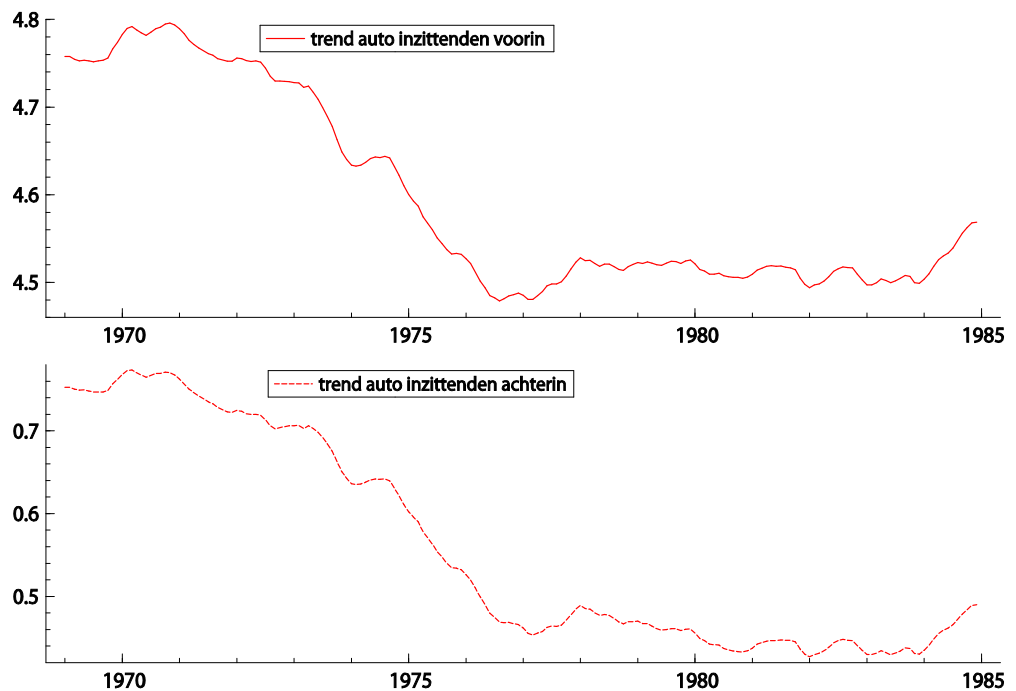
De schattingen van de vier regressiegewichten in model (4) zijn weergegeven in Tabel 11.

	Regressiegewicht	Standaardfout	t-waarde	p-waarde
voor				
log(benzineprijs)	-0.3071	0.1067	-2.8791	0.0045
gordelplicht	-0.3372	0.0495	-6.8157	1.2e-010
achter				
log(benzineprijs)	-0.0836	0.1123	-0.7443	0.4576
gordelplicht	0.0021	0.0518	0.0405	0.9677

Tabel 11. Effecten van benzineprijs en invoering van de gordelplicht op het aantal overleden en zwaargewonde passagiers voor- en achterin de auto.

Deze resultaten impliceren een zeer significante $100(\exp(-0.3372) - 1) = -28.6\%$

reductie van het aantal overleden en zwaargewonde inzittenden voorin de auto door de invoering van de gordelplicht, terwijl het effect van de maatregel op de inzittenden achterin de auto totaal niet significant is.



Figuur 11. Trends van de tijdreeksen in Figuur 6.9.

De met deze analyse verkregen van maandschommelingen ontdane trends zijn weergegeven in Figuur 11. Opvallend aan deze twee trends is dat ze dezelfde variatie over de tijd lijken te vertonen, hetgeen wordt bevestigd door de hoge correlatie van 0.9827 tussen de twee trends.

We herhalen de analyse van de twee tijdreeksen dan ook maar nu zonder interventie variabele op de reeks met inzittenden achterin de auto, en met toevoeging van de restrictie dat de twee trends perfect met elkaar moeten correleren. Deze laatste analyse resulteert in de regressiegewichten van Tabel 12.

	Regressiegewicht	Standaardfout	t-waarde	p-waarde
voor				
log(benzineprijs)	-0.3153	0.1037	-3.0410	0.0027
gordelplicht	-0.3387	0.0196	-17.2850	1.1e-040
achter				
log(benzineprijs)	-0.0819	0.1070	-0.7649	0.4453

Tabel 12. *Effecten van benzineprijs en invoering van de gordelplicht op het aantal overleden en zwaargewonde passagiers voorin de auto, en effect van benzineprijs op het aantal overleden en zwaargewonde passagiers achterin de auto.*

Bij deze analyse blijkt weer dat de invoering van de gordelplicht geassocieerd is met een zeer significante verandering van $100(\exp(-0.3387) - 1) = -28.7\%$ van het aantal overleden en zwaargewonde passagiers voorin de auto. Opmerkelijk is verder dat de waarde van de *t*-toets voor het regressiegewicht van de invoering van de gordelplicht bij inzittenden voorin de auto ten opzichte van de vorige analyse nu met een factor van meer dan 2.5 is gestegen van -6.8157 naar -17.2850 (vergelijk Tabel 11 en 12). Dat komt omdat de standaardfout voor de toets veel kleiner is geworden, en dit is weer het gevolg van de gemeenschappelijke trend die in de laatste analyse aan beide tijdreeksen is opgelegd. Uit de tijdreeks literatuur is bekend dat dit verschijnsel (dat ook wel **coïntegratie** wordt genoemd) kan leiden tot een hogere precisie van het effect van een verklarende variabele, zoals hier ook duidelijk is gebeurd.

We eindigen deze bespreking van de mogelijkheden van ex post evaluatie met tijdreeks analyse met op te merken dat er aan het gebruik van de overleden en zwaargewonde passagiers achterin de auto als referentiegroep ook bezwaren kunnen kleven:

- passagiers achterin de auto zouden meer ernstig letsel kunnen oplopen omdat de inzittenden voorin een gordel dragen en dus in hun stoel blijven zitten bij een botsing;
- passagiers zijn mogelijk vaker achterin de auto gaan zitten of geplaatst om zo geen gordel te hoeven te dragen;
- meer passagiers achterin de auto zijn mogelijk ook een gordel gaan dragen.

In deze studie zijn er echter geen statistische aanwijzingen gevonden dat de gordelplicht enig effect (positief dan wel negatief) heeft gehad op passagiers achterin de auto.

Een ander bezwaar is dat er mogelijke afhankelijkheden zijn tussen passagiers voor- en achterin de auto omdat ze zich in dezelfde auto's verplaatsen. Dit zou ondervangen kunnen worden door ook andere referentiegroepen in het onderzoek te betrekken, zoals het aantal overleden en zwaargewonde passagiers voor- of achterin de auto in een streek waar de gordelplicht (nog) niet is ingevoerd. Ook zouden de volgende twee tijdreeksen met elkaar kunnen worden vergeleken: het aantal overleden en zwaargewonde passagiers voorin de auto als gevolg van ongevallen op even dagen als experimentele groep, en het aantal overleden en zwaargewonde passagiers achterin de auto als gevolg van ongevallen op oneven dagen als referentiegroep.

7. De Empirical Bayes methode

In deze paragraaf bespreken we een tweede manier om ex post evaluaties uit te voeren die - althans in de wereld van verkeersveiligheidsonderzoek - steeds meer opgeld doet en met name ontwikkeld is door Ezra Hauer: de zogenoemde Empirical Bayes (EB) methode. De hier beschreven aanpak is gebaseerd op Hauer (1997). In de EB methode wordt met name bijzonder veel aandacht besteed aan het feit dat de observatie eenheden waarop de onderzochte maatregel wordt toegepast in observatie studies vaak geen aselechte steekproef uit de doelpopulatie vormen maar juist voor behandeling gekozen worden *omdat ze op dat moment relatief erg onveilig zijn (in termen van aantallen ongevallen en/of slachtoffers)*. Aangezien er dan een duidelijke relatie bestaat tussen het aantal ongevallen en/of slachtoffers van de studie objecten en de reden dat juist deze studie objecten aan de maatregel worden onderworpen is er sprake van **selectie bias** hetgeen aanleiding geeft tot het reeds in paragraaf 4 gemelde **regressie naar het gemiddelde effect**. Daarnaast benadrukt ook Hauer dat in een naïeve voor- en nastudie het gevonden effect niet alleen het gevolg hoeft te zijn van de onderzochte maatregel, maar ook van versturende variabelen als verkeersintensiteit, het weer, de omvang van het voertuigpark, het gedrag van bestuurders, de registratiegraad van ongevallen, etc., etc.

Als studie objecten (zoals bepaalde locaties) gekozen zijn vanwege hun hoge onveiligheid dan stelt Hauer voor om te corrigeren voor het regressie naar het gemiddelde effect door altijd *twee* soorten informatie te betrekken bij het bepalen van het verwachte aantal ongevallen van de studie objecten:

1. K , het geobserveerde aantal ongevallen op een onderzochte locatie en
2. $E[\kappa]$, het aantal ongevallen in een **referentie populatie** van locaties met **dezelfde eigenschappen** als de onderzochte locatie.

Het verwachte aantal ongevallen van een studie object wordt berekend middels

$$E[\kappa | K] = \alpha E[\kappa] + (1 - \alpha)K, \quad (5)$$

met $0 \leq \alpha \leq 1$. Hierin is $E[\kappa|K]$ het verwachte aantal ongevallen op de locatie, gegeven het aantal geobserveerde ongevallen op die locatie. Dit verwachte aantal wordt geschat op grond van twee bronnen van informatie: 1) de eigenschappen/kenmerken van het studie object (geslacht, leeftijd, expositie, infrastructuur), en 2) de historie van het aantal ongevallen/slachtoffers. Als α (bijna) 1 is, dan is het aantal verwachte ongevallen op een locatie (vrijwel) gelijk aan het gemiddelde aantal ongevallen in de referentie populatie $E[\kappa]$; als α (bijna) 0 is dan is het aantal verwachte ongevallen op een locatie (vrijwel) gelijk aan het aantal geobserveerde ongevallen K .

De waarde van α in (8.1) wordt bepaald middels

$$\alpha = \frac{1}{1 + \frac{\text{VAR}[\kappa]}{E[\kappa]}}. \quad (6)$$

Hierin zijn $E[\kappa]$ en $VAR[\kappa]$ het gemiddelde en de variantie van het aantal ongevallen in de referentie populatie.

We illustreren de EB procedure aan de hand van het volgende voorbeeld. Stel we hebben een onbewaakte spoorwegovergang buiten de bebouwde kom (bubeko) met een enkel spoor, met 2 treinen per dag en 550 auto's per dag. In vijf jaar tijd, van 1981 t/m 1985, hebben er twee ongevallen plaatsgevonden. Wat is de schatting van het verwachte aantal ongevallen op die locatie?

Uit een ander onderzoek naar de referentiepopulatie van onbewaakte spoorwegovergangen bubeko met enkel spoor is gebleken dat het volgende model

$$\hat{E}[\kappa] = a(\text{auto's / dag})^b (\text{treinen / dag})^c (\text{treinen / dag})^{d \ln(\text{treinen / dag})}$$

leidt tot de volgende parameter waarden:

$$\hat{E}[\kappa] = 0.000954(\text{auto's / dag})^{0.405} (\text{treinen / dag})^{1.039} (\text{treinen / dag})^{-0.115 \ln(\text{treinen / dag})}$$

Merk op dat het verwachte aantal ongevallen $E[\kappa]$ dus niets anders is dan de door het model voorspelde aantal ongevallen (gebaseerd op het aantal *geobserveerde* ongevallen K in de referentie populatie). Gezien diens kenmerken (zie boven) volgt uit de referentie populatie voor de hier onderzochte spoorwegovergang dat

$$\hat{E}[\kappa] = (0.000954)(550)^{0.405} (2)^{1.039} (2)^{-0.115 \ln 2} = 0.0239$$

ongevallen per jaar.

Voor diezelfde referentie populatie worden verder de $VAR[\kappa]$ berekend middels

$$\hat{VAR}[\kappa] = (K - \hat{E}[\kappa])^2 - \hat{E}[\kappa],$$

waarna diezelfde $VAR[\kappa]$ worden voorspeld op grond van dezelfde covariaten als waarmee de $E[\kappa]$ zijn bepaald. Soms, zoals in dit voorbeeld, is het mogelijk om de $VAR[\kappa]$ uit te drukken als een simpele functie van de $E[\kappa]$. In dit specifieke voorbeeld blijkt namelijk dat

$$\hat{VAR}[\kappa] = \hat{E}[\kappa]^2 / 0.52,$$

oftewel dat $VAR[\kappa]$ een simpele kwadratische functie is van $E[\kappa]$. Losgelaten op dit voorbeeld is dus

$$\hat{VAR}[\kappa] = \hat{E}[\kappa]^2 / 0.52 = 0.0239^2 / 0.52 = 0.0011$$

ongevallen per jaar. Gedurende vijf jaar is dus

$$\hat{E}[\kappa] = (5)(0.0239) = 0.1195, \quad \hat{VAR}[\kappa] = (5^2)(0.0011) = 0.0275,$$

en

$$\alpha = \frac{1}{1 + \frac{0.0275}{0.1195}} = 0.81,$$

waarna

$$E[\kappa | 2] = \alpha E[\kappa] + (1 - \alpha)(2) = (0.81)(0.1195) + (1 - 0.81)(2) = 0.48$$

ongevallen in vijf jaar.

Hiermee is een inschatting verkregen voor het "werkelijke" aantal ongevallen in de vijf jaar voormeting van deze specifieke spoorwegovergang. Gecorrigeerd voor regressie naar het gemiddelde verwachten we volgens de EB methode bij deze specifieke spoorwegovergang dus geen 2 maar 0.48 ongevallen in vijf jaar. Dit aantal wordt vervolgens gebruikt als inschatting van het aantal ongevallen in de volgende vijf jaar *indien de maatregel niet zou zijn ingevoerd*. Dit laatste aantal wordt vervolgens vergeleken met het *geobserveerde* aantal ongevallen in de naperiode om tot een inschatting te komen van het effect van de maatregel (wat die maatregel ook moge zijn geweest).

Samenvattend bestaat de EB methode voor de ex post evaluatie van de effecten van verkeersveiligheidsmaatregelen uit de volgende stappen:

1. observeer het aantal ongevallen voor de eenheden waarop de maatregel is toegepast (of een steekproef daaruit) in de voor- en naperiode;
2. bepaal het verwachte aantal ongevallen voor diezelfde eenheden in de voorperiode door het aantal ongevallen en de variatie daarin in een referentie populatie te bepalen;
3. corrigeer het geobserveerde aantal ongevallen in de voorperiode voor het verwachte aantal ongevallen in de referentie groep;
4. vergelijk dit laatste gecorrigeerde aantal ongevallen uit de voorperiode met het geobserveerde aantal ongevallen in de naperiode, en baseer op *dit verschil* het effect van de maatregel.

Met deze aanpak wordt getracht zoveel mogelijk te corrigeren voor het regressie naar het gemiddelde effect dat optreedt wanneer bij een observationele studie het vermoeden bestaat dat er een samenhang is tussen de maatregel zelf en de reden waarom de maatregel juist bij deze onderzoekseenheden wordt onderzocht en toegepast.

Voorwaarde voor de toepasbaarheid van de EB methode is uiteraard dat er schattingen beschikbaar moeten zijn over de verwachte waarde van het studieobject middels onderzoek naar de samenhang tussen ongevallen en/of slachtoffers en de eigenschappen van objecten in een referentie populatie.

8. Conclusies

Duidelijk is dat de bewijsvoering rond de ex post evaluatie van verkeersveiligheidsmaatregelen in een natuurlijke setting nooit helemaal waterdicht zal kunnen zijn, met name in termen van causaliteit. We hopen echter wel te hebben laten zien hoe de bewijsvoering rond naïeve voor- en nastudies aanzienlijk sluitender en overtuigender kan worden gemaakt door - waar mogelijk:

1. langere tijdreeksen van observaties in de voorperiode te gebruiken om te kunnen corrigeren voor algemene trends en regressie naar het gemiddelde effecten;
2. voor versturende variabelen die bekend én tijdens voor- en nameting gemeten zijn expliciet te corrigeren door ze als covariaten mee te nemen in de interventie analyse;
3. voor versturende variabelen die onbekend en/of niet gemeten zijn impliciet te corrigeren door een referentiegroep bij de analyse te betrekken;
4. verschillende soorten analyse te gebruiken om effectschattingen van een maatregel te kunnen kruisvalideren.

Bijlage 2: Beleidsdocumenten ex post vragen

1. Strategisch plan verkeersveiligheid: Hierin wordt, onder het kopje "kennisontwikkeling" de behoefte aan kennis over werkelijke effecten van maatregelen (ex post) uitgesproken, met name bij decentrale overheden. Ook wordt vermeld dat weinig bekend is over de effecten van gedragsmaatregelen. In het Actieprogramma zal worden aangegeven wie deze kennisontwikkeling oppakt.
2. Actieprogramma Verkeersveiligheid: aangegeven wordt dat aanvullende mogelijkheden worden onderzocht om de relatie verkeersveiligheid/kosteneffectiviteit in beeld te brengen; hierbij worden de kennisinstellingen nadrukkelijk betrokken.
3. Toezegging AO 2 september 2009: De Minister zegt op basis van een discussie over de effectiviteit van maatregelen toe dat de Kamer schriftelijk geïnformeerd zal worden over de mogelijkheid om de kosteneffectiviteit van maatregelen in beeld te brengen.
4. Brief Minister op 21 oktober 2009 n.a.v. AO 2 september: De Minister meldt: "...dat ik het effect van de beleidsmaatregelen gericht op het verbeteren van de verkeersveiligheid ga onderzoeken. Bij de keuze voor maatregelen vormde proportionaliteit een belangrijk ijkpunt. Aldus ook uw Kamer, die instemde met het Strategisch Plan Verkeersveiligheid. Er zijn natuurlijk maatregelen te bedenken die een vergaand effect op verkeersveiligheid hebben maar zo drastisch ingrijpen in de samenleving dat ik daar nu niet voor gekozen heb. Daarmee wil ik aangeven dat sommige maatregelen wellicht zeer kosteneffectief zouden kunnen zijn (i.e. weinig geld kosten en veel doden kunnen besparen) maar die om bovengenoemde reden toch niet genomen zijn. Daarnaast wil ik nadrukkelijk aangeven dat, naast het kostenaspect, voor mij vooral meetelt hoeveel slachtoffers er met een maatregel worden bespaard. Deze afweging maak ik op basis van een (ex ante) inschatting vooraf. Dit bovengenoemde onderzoek zal zich dan ook in eerste instantie moeten richten op het (ex post) evalueren van het daadwerkelijk behaalde effect in termen van bespaarde slachtoffers. Een indicatie van de kosten van maatregelen is daarna wellicht ook mogelijk, maar ik herhaal daarbij dat het effect van maatregelen (en dus een directe toerekening van kosten van maatregelen aan bespaarde verkeersdoden) moeilijk specifiek meetbaar is. De effecten van maatregelen komen voort uit een mix van onder andere educatie, regelgeving, communicatie en handhaving, waardoor het lastig is om effecten specifiek aan een van deze onderdelen toe te rekenen. Ik heb Rijkswaterstaat gevraagd om samen met onder andere de SWOV te onderzoeken of het mogelijk is om met een methode te komen die de van tevoren ingeschatte besparing van slachtoffers achteraf kan evalueren."