

Wanneer welke evaluatie-opzet: Kieswijzer voor interventies in de publieke gezondheid

S.C. Jansen¹, A. Haveman-Nies², P. van 't Veer²

De evaluatie van interventies heeft een belangrijke plaats in de publieke gezondheid. Zowel praktijk, beleid als wetenschap zijn gebaat bij een antwoord op de vraag: werkt de interventie? Welke evaluatie-opzet gebruikt moet worden om deze vraag te beantwoorden, hangt af van het type interventie en de resultaten die men wil bereiken. Wij zijn ervan overtuigd dat zowel de evaluatie-opzet zonder controlegroep, de evaluatie-opzet met controlegroep en de randomised controlled trial (rct) bestaansrecht hebben in de publieke gezondheid. Om de juiste evaluatie-opzet te kunnen kiezen, zijn twee zaken van doorslaggevend belang. Ten eerste het niveau van resultaten dat men wil gaan meten: dichtbij of veraf van de interventie in tijd en causale keten. Voor resultaten dichtbij de interventie volstaat de evaluatie-opzet zonder controlegroep, voor resultaten verder weg is een controlegroep noodzakelijk. De tweede doorslaggevende factor is de zekerheid waarmee men de vraag: zijn de resultaten toe te schrijven aan de interventie? beantwoord wil zien. De *gewenste* zekerheid van conclusies bepaalt de keuze tussen de evaluatie-opzet met controlegroep en de rct. Op basis van deze principes hebben wij de Kieswijzer ontwikkeld. Deze Kieswijzer faciliteert de keuze van de evaluatie-opzet voor interventies in de publieke gezondheid.

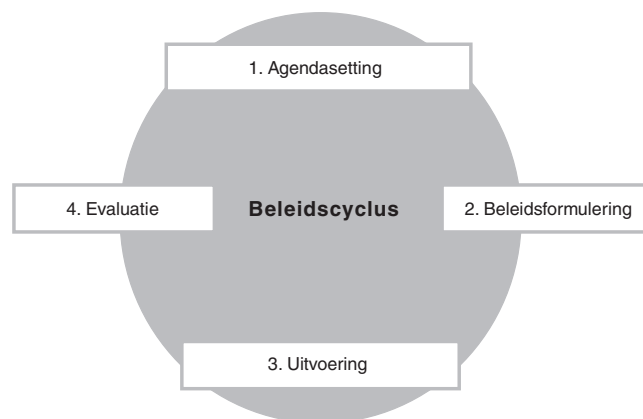
Trefwoorden: evaluatie, evaluatie-opzet, interne validiteit, interventies, publieke gezondheid, gezondheidsbevordering

INLEIDING

De evaluatie van interventies heeft een belangrijke plaats in de publieke gezondheid. Dit is te verduidelijken met de beleidscyclus, een werkmodel dat veel gehanteerd wordt in de publieke gezondheid (figuur 1). Evaluatie is hierin één van de fasen, naast agendasetting, beleidsformulering en uitvoering. In de evaluatie-fase staat de vraag centraal: hebben we onze doelen bereikt? Omdat (lokale) overheden verantwoording moeten afleggen over hun beleid, wordt steeds vaker gevraagd naar de resultaten van interventies. Dat evaluatie-onderzoek steeds belangrijker wordt, blijkt ook uit de recente oprichting van de werkplaats certificering van het Centrum Gezond Leven.¹ Interventies waarvan de werkzaamheid is aangetoond komen in aanmerking voor certificering. Op dit moment staan wij voor de uitdaging om de vele interventies, die vanuit lokaal initiatief worden ontwikkeld en uitgevoerd, te evalueren. Ten behoeve van dit artikel definiëren wij interventies als een (pakket van) actie(s) met een coherent doel, gericht op het veroorzaken van verandering. Hieronder vallen bijvoorbeeld educatie en voorlichting, maar ook o.a. beleidsmaatregelen, lobby-activiteiten en

sociale activering. Interventies in de publieke gezondheid zijn gericht op bevorderen of beschermen van de gezondheid of op het voorkómen van ziekten in gemeenschappen of populaties.²

Wanneer kan worden geconcludeerd dat een interventie werkt? Ten eerste moet vastgesteld worden of het resultaat, dat met de interventie beoogd werd, bereikt is. Bijvoorbeeld: is de samenwerking tussen organisaties verbeterd, zijn jongeren minder alcohol gaan drinken, is



Figuur 1 De beleidscyclus

¹ GGD Gelre-IJssel / Academische werkplaats Agora

² Wageningen Universiteit, afdeling Humane Voeding / Academische werkplaats Agora

de buikomvang afgenomen, zijn er meer groene zones in de stad, et cetera. Ten tweede moet de vraag: is dit resultaat toe te schrijven aan de interventie? gesteld en beantwoord worden. In meer populaire zin uitgedrukt: komt het echt door de interventie? Dat is geen puur wetenschappelijke vraag, maar een vraag die relevant is voor de praktijk van de publieke gezondheid. Niet elke verbetering is immers automatisch het gevolg van de interventie. Het resultaat dat gemeten wordt, kan bijvoorbeeld vertekend zijn door landelijke campagnes of gevolg zijn van toeval.

De evaluatie-opzet dient ervoor om de vraag: zijn de resultaten toe te schrijven aan de interventie? te beantwoorden. Dat maakt de evaluatie-opzet onmisbaar bij het evalueren van interventies. Op dit moment ontbreekt echter een 'best practice' voor de keuze van de evaluatie-opzet. Er wordt al jarenlang gediscussieerd over de juiste evaluatie-opzet voor interventies in de publieke gezondheid, zowel in Nederland^{3,4,5} als internationaal.^{6,7,8} Tot nu toe lijkt het veld verdeeld te zijn in voor- en tegenstanders van één van de varianten van de evaluatie-opzet, de randomised controlled trial (rct). Voorstanders zien de rct als gouden standaard, omdat deze het meest zekere antwoord geeft op de vraag of de resultaten zijn toe te schrijven aan de interventie. Tegenstanders daarentegen verwerpen de rct rigoureuus omdat deze strijdig is met de principes van gezondheidsbevordering zoals beschreven in de Ottawa Charter,⁹ bijvoorbeeld het combineren van strategieën en het samenwerken met lokale partijen bij de ontwikkeling en uitvoering van de interventie. Zij stellen voor om interventies te evalueren met participatieve en theorie-gebaseerde evaluatie-methoden.^{10,11} Hoewel dit waardevolle methoden zijn, lijkt hiermee echter de stap van de keuze van de evaluatie-opzet, die hieraan vooraf gaat, te worden overgeslagen.

Het valt ons op dat aan de twee uitersten van de rct-discussie heel anders wordt omgegaan met de vraag: zijn de resultaten toe te schrijven aan de interventie? Voorstanders van de rct lijken er vanuit te gaan, dat deze vraag altijd zo zeker mogelijk beantwoord moet worden, terwijl deze vraag bij tegenstanders van de rct nauwelijks een rol lijkt te spelen.

DOEL

In dit opiniërend artikel staat de evaluatie-opzet centraal als onmisbare basis voor de evaluatie van interventies in de publieke gezondheid. We pleiten ervoor dat de vraag: zijn de resultaten toe te schrijven aan de inter-

Kernpunten

- Evaluatie van interventies in de publieke gezondheid wordt steeds belangrijker; de evaluatie-opzet is hierbij een onmisbare basis.
- De evaluatie-opzet bepaalt de zekerheid van conclusies over de resultaten: zijn deze echt door de interventie veroorzaakt en niet door andere factoren?
- Zowel de evaluatie-opzet zonder controlegroep, de evaluatie-opzet met controlegroep als de randomised controlled trial (rct) hebben bestaansrecht in de publieke gezondheid. Wanneer welke evaluatie-opzet moet worden gekozen hangt af van het type interventie en de resultaten die men wil bereiken.
- De Kieswijzer faciliteert de keuze van een passende evaluatie-opzet voor interventies op het grensvlak van praktijk, beleid en wetenschap in de publieke gezondheid.

ventie? vaker wordt gesteld en beantwoord. Tegelijkertijd willen wij de vanzelfsprekendheid waarmee wordt aangenomen dat deze vraag altijd zo zeker mogelijk beantwoord dient te worden, ter discussie stellen. We gaan in op de hoofdvarianten van de evaluatie-opzet en de zekerheid van conclusies die daaraan verbonden mogen worden. Vervolgens geven wij onze visie op de evaluatie-opzet in de publieke gezondheid. Wij presenteren de Kieswijzer om de keuze van de evaluatie-opzet in de praktijk te faciliteren. Hiermee hopen wij bij te dragen aan de totstandkoming van een 'best practice' voor kwalitatief goed evaluatie-onderzoek in de praktijk.

DE EVALUATIE-OPZET EN HAAR HOOFDARIANTEN

De evaluatie-opzet (synoniemen: evaluatie-ontwerp, study design) is de *structuur* die specifieke details van de te meten groep(en), tijdsbestek, methode en procedures van evaluatie bevat, inclusief ethische overwegingen (vrij vertaald naar de definitie van research design¹²). De evaluatie-opzet is een overkoepelend geheel, waarbinnen evaluatie-methoden een onderdeel zijn. Bij de keuze van de evaluatie-opzet wordt gedefinieerd *wanneer, bij wie en hoe* er wordt gemeten. *Wanneer*: wordt er alleen na afloop van de interventie gemeten of ook vooraf, en wordt er ook tussentijds gemeten? En *bij wie*: worden alle mensen gemeten die men met de interventie wil bereiken, of een gedeelte (steekproef)? Hoe wordt deze steekproef geselecteerd en moet deze representatief zijn voor de totale doelgroep? Wordt er ook gemeten in een controlegroep of niet? Ook wordt in de evaluatie-opzet bepaald *hoe* er wordt gemeten: welke methode en procedures worden gebruikt? Heeft de onderzoeker invloed op de omstandigheden? Maakt hij deel uit van het onderzoek (zoals bij participatief onderzoek) of staat hij meer op afstand? En, afhankelijk van de aard van de resultaten: worden kwalitatieve of kwantitatieve methoden gebruikt om de resultaten te meten? Met methoden bedoelen wij bijvoorbeeld lichamelijke metingen, schriftelijke vragenlijsten, registraties, interviews, focusgroepen, observaties en dossieranalyse. Evaluatie-methoden moeten niet verward worden met de evaluatie-opzet; deze zijn van een andere orde.

Op hoofdlijnen zijn er drie varianten van de evaluatie-opzet te onderscheiden:^{13,14}

- 1 de evaluatie-opzet zonder controlegroep (synoniemen: pre-experimentele opzet, non-experimental design)
- 2 de evaluatie-opzet met controlegroep (synoniemen: quasi-experimentele opzet, non-equivalent control group design, non-randomised design)

3 de rct/cit (synoniemen: experimentele opzet, experimenteel design)

In elke variant vinden één of meerdere voormetingen, tussenmetingen en/of nametingen plaats. Bij variant (1) vinden deze metingen alleen plaats in de interventiegroep; bij variant (2) en (3) ook in een controlegroep. Bij de rct/cit (variant 3) is er bovendien sprake van *randomisatie*: de interventie wordt willekeurig toegewezen aan individuen, waardoor een identieke interventiegroep en controlegroep ontstaan. De community intervention trial (cit) is een variant van de rct waarbij de randomisatie-eenheid niet bestaat uit individuen maar uit groepen, zoals huisartsenpraktijken, scholen of wijken. Naast randomisatie is bij de rct/cit ook sprake van *standaardisatie* van zowel interventie als omstandigheden: de omstandigheden worden zoveel mogelijk onder controle gehouden en de interventie wordt in elk individu of groep volgens identiek protocol uitgevoerd. Het woord *controlled* in 'randomised controlled trial' betekent dus niet alleen dat er een controlegroep aanwezig is, maar ook dat er zoveel mogelijk controle wordt uitgeoefend op de interventie en omstandigheden.¹⁵ Overigens wordt de rct ook wel 'randomised clinical trial' genoemd, omdat ze in de klinisch-wetenschappelijke traditie veel wordt gebruikt.

In tabel 1 worden de kenmerken van de drie hoofdvarianten van de evaluatie-opzet op een rij gezet. In onze visie is evalueren zonder evaluatie-opzet onmogelijk: bij elke evaluatie van een interventie in de publieke gezondheid is sprake van één van de drie hoofdvarianten. Als de evaluatie-opzet niet expliciet wordt benoemd, is er meestal sprake van een evaluatie-opzet zonder controlegroep. Veel participatieve evaluatie-methoden, zoals actiebegeleidend onderzoek, worden toegepast binnen een evaluatie-opzet zonder controlegroep. Het is belangrijk om de evaluatie-opzet expliciet te benoemen, omdat deze van invloed is op de conclusies die na afloop van de interventie getrokken mogen worden.

DE EVALUATIE-OPZET EN DE ZEKERHEID VAN CONCLUSIES

De vraag: zijn de resultaten toe te schrijven aan de interventie? kan niet met elke variant van de evaluatie-opzet even goed beantwoord worden. Allerlei verstoringen (ook wel genoemd: bedreigingen van de interne validiteit) kunnen het resultaat, dat na afloop van de interventie wordt gemeten, hebben beïnvloed. Zoals bijvoorbeeld vergelijkbare interventies die door anderen zijn geïnitieerd en in dezelfde periode plaatsvinden. Of het ouder worden van de doelgroep, een landelijke trend, natuurlijke schommeling, selectieve uitval van deelnemers etcetera. Met de evaluatie-opzet kunnen deze ver-

storende factoren worden uitgeschakeld. De elementen van de evaluatie-opzet: voormeting, nameting, controlegroep, randomisatie en standaardisatie hebben hierin ieder een eigen rol. Naarmate er meer van deze elementen in de evaluatie-opzet worden gebruikt, kunnen verstoringen beter worden uitgeschakeld. En is dus een zekerder antwoord mogelijk op de vraag: zijn de resultaten toe te schrijven aan de interventie?

Niet elke variant van de evaluatie-opzet kan dus hetzelfde antwoord geven op de vraag: zijn de resultaten toe te schrijven aan de interventie? De antwoorden variëren als volgt:^{13,14,16,17}

- 1 Met een **evaluatie-opzet zonder controlegroep** moet men het antwoord eigenlijk schuldig blijven: er is teveel kans dat het resultaat door andere factoren veroorzaakt is.
- 2 Een **evaluatie-opzet met controlegroep** maakt het aannemelijk (maar nooit helemaal zeker), dat het resultaat aan de interventie is toe te schrijven.
- 3 Met een **rct/cit** kan stellig worden geconcludeerd: de resultaten zijn het gevolg van de interventie.

De zekerheid van het antwoord op de vraag: zijn de resultaten toe te schrijven aan de interventie? wordt dus in de eerste plaats bepaald door de hoofdvariant van de evaluatie-opzet. In de tweede plaats wordt deze zekerheid bepaald door de specifieke invulling van de hoofdvariant, zoals de omvang van de te meten groep, het aantal tussenmetingen, de gekozen meetmethode etcetera. Wanneer bijvoorbeeld de evaluatie-opzet zonder controlegroep wordt ingevuld op de manier van een *time-series design*, neemt de zekerheid van conclusies sterk toe. In een *time-series design* wordt namelijk met vele metingen vooraf, tijdens en na afloop van de interventie een trend in de tijd zichtbaar gemaakt. Wanneer een trendbreuk zichtbaar is direct nadat de interventie werd uitgevoerd, is dit een aanwijzing dat dit resultaat door de interventie veroorzaakt is. Dit kan een even zekere conclusie geven als een evaluatie-opzet met controlegroep. Binnen de hoofdvariant rct/cit levert de cit een minder zekere conclusie, omdat randomisatie en standaardisatie in de cit moeilijk te realiseren zijn. De drie hoofdvarianten van de evaluatie-opzet kunnen elkaar dus gedeeltelijk overlappen in de zekerheid van conclusies.

De rct/cit geeft de meest zekere conclusie. Echter, bij de uitvoer van de rct/cit in de praktijk van de publieke gezondheid stuit men op allerlei problemen, die met name te maken hebben met randomisatie en standaardisatie.^{18,19} Het random toewijzen van een interventie aan individuen is vaak onmogelijk, zoals bij mediacampagnes en interventies met een wijk-aanpak (community-interventies), die per definitie aan de gehele gemeenschap worden aangeboden. Bovendien is het random toewijzen

Tabel 1 De drie hoofdvarianten van de evaluatie-opzet en hun kenmerkende elementen

	voormeting tussenmeting en/of nameting	controlegroep	randomisatie	standaardisatie
1 evaluatie-opzet zonder controlegroep	x			
2 evaluatie-opzet met controlegroep	x	x		
3 rct / cit	x	x	x	x

van een interventie aan groepen in de praktijk vaak onwenselijk, omdat heel gericht wijken, gemeenten, huisartspraktijken of scholen geselecteerd worden voor het interventieprogramma op basis van aanwezig enthousiasme of bijvoorbeeld de toezegging van een wethouder. Ook het standaardiseren van de interventie en omstandigheden staat op gespannen voet met de levensechte situatie zoals die in de praktijk nu eenmaal is. Het is bijvoorbeeld niet mogelijk om maximale controle op de uitvoering van een interventie te behouden en tegelijkertijd intensief samen te werken met lokale partijen. Randomisatie en standaardisatie ten behoeve van de evaluatie van interventies in de publieke gezondheid is dus soms onmogelijk en soms onwenselijk.

VOLDOENDE ZEKERHEID OM TE KUNNEN HANDELEN

Ons inziens hoeft bij het evalueren van interventies in de publieke gezondheid niet altijd gestreefd te worden naar een zo zeker mogelijk antwoord op de vraag: zijn de resultaten toe te schrijven aan de interventie? Het gaat er immers niet om, deze vraag zo wetenschappelijk mogelijk te beantwoorden. Het gaat erom, voldoende input te leveren voor de (lokale) besluitvorming binnen beleid en bestuur over het inkopen, continueren, aanpassen of stopzetten van gezondheidsbevorderende en preventie interventies. Voorafgaand aan een evaluatie is het zinvol om te bekijken, wie welke beslissing moet nemen op basis van de evaluatie en welke zekerheid nodig is om te kunnen handelen. Het is niet altijd nodig de top van de wetenschappelijke piramide te bestijgen en een waterdichte conclusie over het effect van de interventie, met een juiste schatting van de effectgrootte, te leveren. Soms is het voldoende, als er redelijke aanwijzingen zijn dat de interventie werkt en geen averechtse effecten heeft. Factoren die een rol spelen bij de gewenste zekerheid van conclusies zijn onder andere: het niveau van besluitvorming (lokaal of nationaal), de kosten van de interventie en de kans op nadelige bijwerkingen. Voor interventies die landelijk worden uitgerold, met hoge kosten of een grote kans op nadelige bijwerkingen zijn zekerder conclusies gewenst dan voor relatief goedkope interventies die op lokaal niveau worden uitgevoerd. Onze stelling is, dat het in de publieke gezondheid meestal voldoende is om *aannemelijk* te maken dat de resultaten het gevolg zijn van de interventie. De evaluatie-opzet met niet-gerandomiseerde controlegroep kan deze conclusie leveren.

DE EVALUATIE-OPZET EN HET NIVEAU VAN RESULTATEN

Tot nu toe hebben wij geredeneerd vanuit het uitgangspunt, dat het resultaat van de interventie beïnvloed kan zijn door versturende factoren. De evaluatie-opzet is dan nodig om deze factoren uit te schakelen. Op dit uitgangspunt zijn diverse effectladders of 'evidence-piramides' gebaseerd, die gebruikt worden om de bewijskracht van een evaluatie-opzet te bepalen.^{16,17} Echter, de vraag is of dit uitgangspunt altijd opgaat – er zijn ook situaties mogelijk waarin het resultaat nauwelijks door versturende factoren beïnvloed kan zijn. Denk bijvoorbeeld aan

interventies die erop gericht zijn om sociaal draagvlak te creëren, samenwerkingsverbanden te realiseren of financieringsstromen te beïnvloeden. Of aan resultaten op procesniveau, zoals bereik en waardering van de interventie. Het is algemeen aanvaard dat dergelijke resultaten worden vastgesteld met een evaluatie-opzet zonder controlegroep. Kennelijk bestaat er een bepaald *niveau van resultaten* waarbij de invloed van versturende factoren zo klein is, dat de resultaten 'van nature' aan de interventie kunnen worden toegeschreven.

De idee dat resultaten van interventies op verschillende niveaus gemeten kunnen worden, vinden we onder andere terug bij Nutbeam.¹⁹ Hij onderscheidt in zijn model vier niveaus van uitkomsten: 0 'health promotion actions', 1 'health promotion outcomes', 2 'intermediate health outcomes' en 3 'health and social outcomes' (figuur 2). Deze figuur moet gelezen worden als een tijdsbalk van onder naar boven: eerst vinden de acties (interventies) plaats, daarna ontstaat verandering in niveau 1, vervolgens in niveau 2 en weer later in niveau 3. De 'health promotion outcomes' liggen het dichtst bij de interventie in tijd: er verstrijkt, in vergelijking met de andere niveaus, weinig tijd totdat hier resultaat te zien is. De 'health promotion outcomes' liggen bovendien het dichtst bij de interventie in causale keten: er zijn, in vergelijking met de andere niveaus, weinig tussenstappen nodig voordat op dit niveau resultaat bereikt wordt. De 'health and social outcomes' daarentegen liggen veraf van de interventie in tijd en causale keten: er zijn heel wat jaren en tussenstappen nodig, gezien vanaf de interventie, voordat hier resultaat wordt bereikt. Saan & de Haes presenteren met hun Referentiekader voor Gezondheidsbevordering een vergelijkbaar model.²⁰ Een verschil is dat zij niet beginnen met het niveau 0 'interventies' (vergelijkbaar met 'health promotion actions' van Nutbeam) maar met -1 'organiseren', wat gaat over de praktische randvoorwaarden voor interventies. Dat betekent een extra niveau waarop resultaat kan worden geboekt. In Box 1 worden de niveaus van resultaten aan de hand van een voorbeeld toegelicht.

CONTROLEGROEP NIET ALTIJD NODIG

De niveaus waarop resultaten van interventies gemeten kunnen worden, verschillen dus in de afstand die zij hebben tot de interventie in tijd en causale keten. Er zijn niveaus 'dichtbij' en niveaus 'veraf' van de interventie. Waar het nu om gaat is het volgende: van resultaten dichtbij de interventie kan *redelijkerwijs beweerd worden* dat zij het gevolg zijn van de interventie. Zij zijn namelijk direct met de interventie verbonden; de kans dat zij beïnvloed zijn door versturende factoren is heel klein. Voor de keuze van de evaluatie-opzet is het niveau van resultaten dat men wil gaan meten dus van groot belang. Als men zich beperkt tot doelen (=gewenste resultaten) die dichtbij de interventie liggen in tijd en causale keten, kan men volstaan met een evaluatie-opzet zonder controlegroep. Wanneer men resultaten wil meten die grotere afstand hebben tot de interventie, is een controlegroep noodzakelijk.

In de praktijk zal men van een specifieke interventie

Box 1: niveaus van resultaten

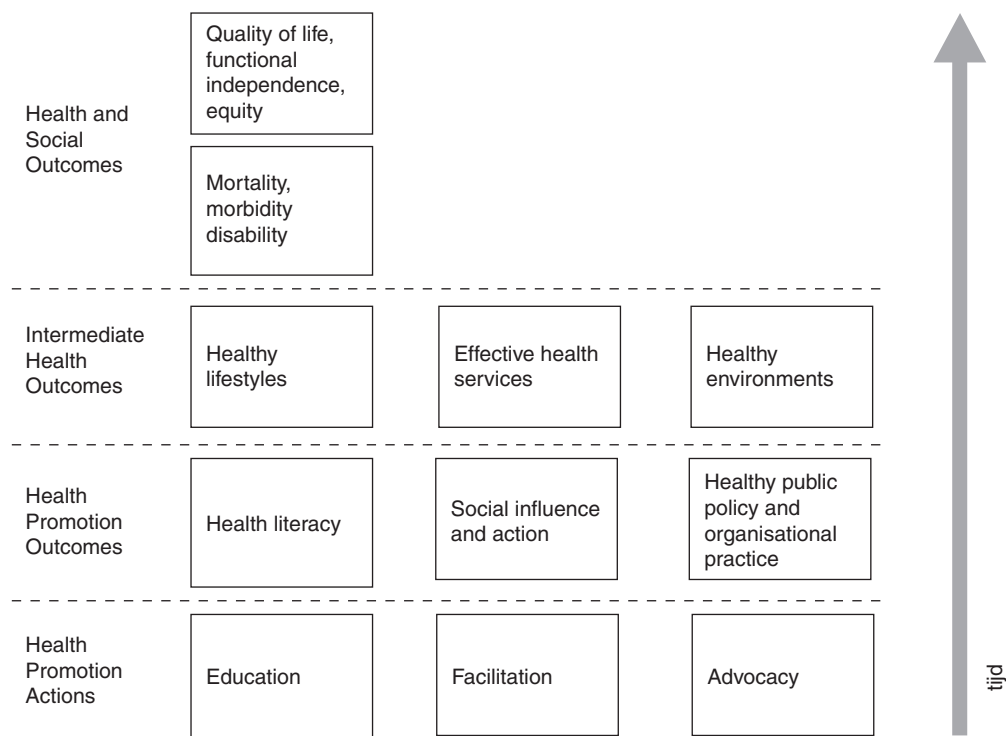
In de regio Achterhoek wordt sinds enkele jaren de complexe interventie: 'Alcoholmatiging jeugd in de Achterhoek' uitgevoerd.^{25,26} Deze interventie is begonnen in de praktijk met het organiseren van de randvoorwaarden, met name de deelname van en financiering door gemeenten. Toen alle Achterhoekse gemeenten te kennen gaven mee te willen doen met het project, was resultaat op dit niveau (- 1) behaald.

Vervolgens werd op basis van de beschikbare theoretische en empirische kennis een interventiepakket ontworpen met een multidisciplinaire aanpak. Dit was een resultaat op niveau 0 'interventies'. Gedurende de interventie werd gemeten of de interventie werd uitgevoerd zoals gepland, hoeveel jongeren en hun ouders zijn bereikt met de verschillende onderdelen van de interventie, en of zij deze konden waarderen: resultaten op het niveau direct na de interventie. De samenwerking met diverse partijen werd besproken middels interviews en waar nodig bijgesteld: resultaten die direct verbonden zijn met de interventie.

Bij ouders werden de kennis over de negatieve gevolgen van alcohol, de opvoedingsvaardigheden en de sociale norm gemeten: deze resultaten liggen al wat verder af van de interventie (niveau 1).²⁶

Het eindresultaat was het drinkgedrag van jongeren zelf: dit ligt veraf van de interventie in tijd en causale keten, namelijk op niveau 2 'determinanten van gezondheid'. Het drinkgedrag werd geëvalueerd door beschikbare data uit een grootschalig vragenlijsten-onderzoek onder scholieren in de gehele regio Oost-Nederland in 2003 en 2008 te analyseren (resultaten nog niet gepubliceerd).

Tenslotte, doelen op niveau 3 'gezondheid' werden in het alcoholmatigingsproject niet gesteld en dus ook niet gemeten, omdat in andere onderzoeken voldoende is aangetoond dat een vermindering van alcoholgebruik gunstige gevolgen heeft voor de gezondheid en de veiligheid van jongeren.



Figuur 2 An outcome model for health promotion;²⁹ pijl toegevoegd door de auteurs.

vaak meerdere (tussen)resultaten willen meten die zowel dichtbij als veraf van de interventie liggen. Het is dan van belang om voor elk (tussen)resultaat te bepalen welke evaluatie-opzet passend is. In het alcoholmatigingsproject (box 1) werd de samenwerking met lokale partijen bijvoorbeeld gemeten met interviews binnen een evaluatie-opzet zonder controlegroep. Het drinkgedrag van de jongeren werd gemeten met schriftelijke vragenlijsten binnen een evaluatie-opzet met controlegroep. Ons inziens volstaat de gebruikelijke indeling in proces- en effectmaten hier niet.

Voor resultaten zoals het ontstaan van een netwerkstructuur, betere handhaving van regels of een veranderde sociale norm is immers moeilijk te zeggen of zij onder proces of effect vallen. Vaak ontstaat hierover spraakverwarring tussen epidemiologen, gezondheidsbevorderaars en beleidsmakers.^{21,22} Ons inziens laten modellen zoals van Nutbeam en Saan & de Haes zien, dat er geen strikte scheiding is tussen proces en effect. In werkelijkheid liggen proces- en effectmaten ergens op de glijdende schaal van resultaten die dichtbij tot

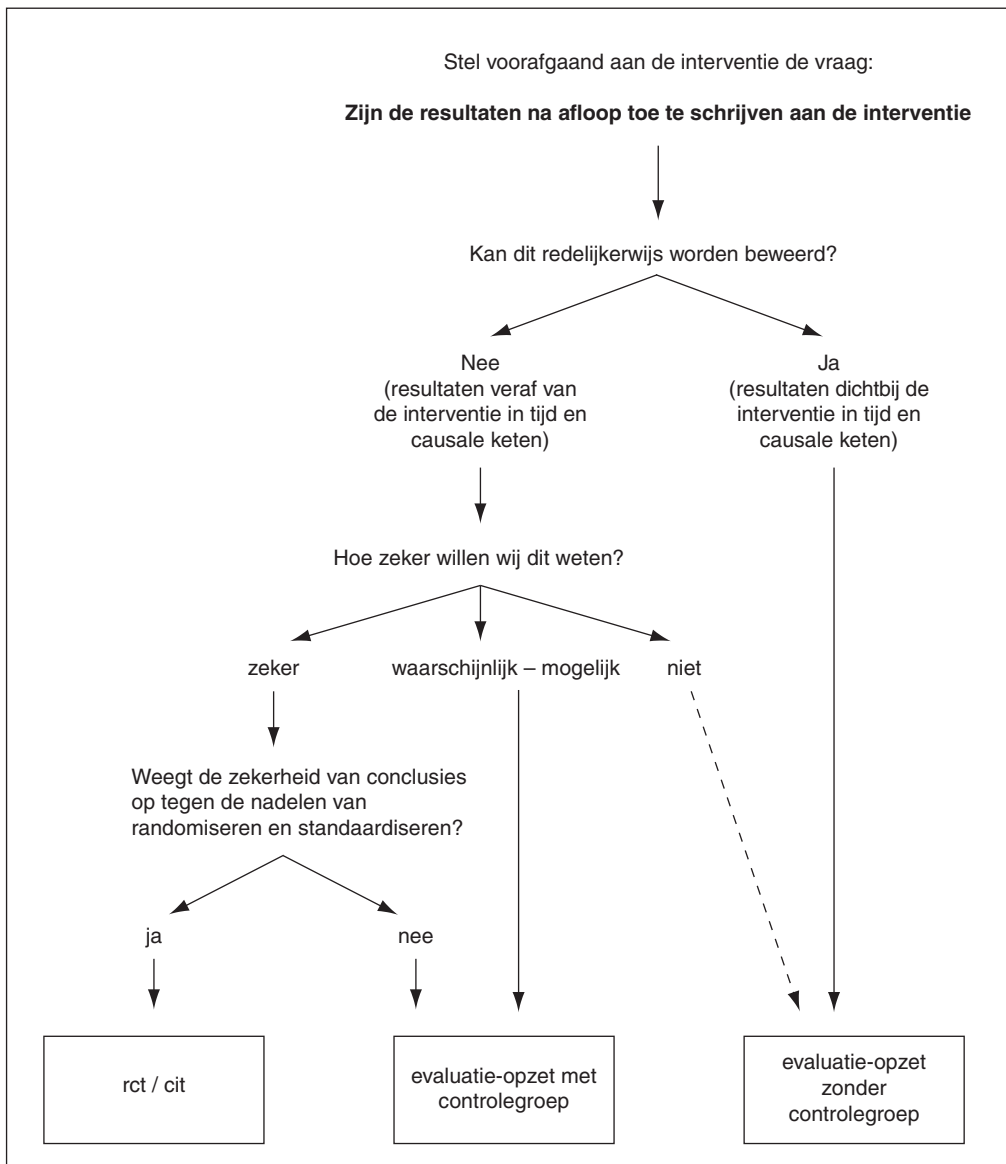
ver weg van de interventie liggen in tijd en causale keten. Het is dus belangrijk om de logische tussenstappen in beeld te hebben. Wanneer verbetering optreedt in diverse tussenstappen van de causale keten, is dit een sterke aanwijzing dat de interventie werkt.

KIESWIJZER VOOR EVALUATIE-OPZET

In figuur 3 presenteren wij de Kieswijzer. Deze Kieswijzer faciliteert de keuze voor een passende evaluatie-opzet, gebaseerd op de principes die wij hierboven uiteen hebben gezet. De vraag: zijn de resultaten toe te schrijven aan de interventie? staat hierbij centraal. Om de juiste evaluatie-opzet te kunnen kiezen, zijn twee zaken van doorslaggevend belang. Ten eerste het niveau van resultaten dat men wil gaan meten: dichtbij of veraf van de interventie in tijd en causale keten. Voor resultaten dichtbij de interventie volstaat de evaluatie-opzet zonder controlegroep, voor resultaten verder weg is een controlegroep noodzakelijk. De tweede doorslaggevende factor is de zekerheid waarmee men de vraag: zijn de resultaten toe te

schrijven aan de interventie? beantwoord wil zien. De *gewenste* zekerheid van conclusies bepaalt de keuze tussen de evaluatie-opzet met controlegroep en de rct/cit. Ons inziens moet de rct/cit alleen worden ingezet als de extra zekerheid van conclusies opweegt tegen de nadelen van het randomiseren en standaardiseren. In sommige gevallen zal randomisatie en/of standaardisatie onmogelijk zijn, terwijl toch een zo zeker mogelijke conclusie gewenst is. Dan is het raadzaam om te zoeken naar alternatieven voor randomisatie en/of standaardisatie, bijvoorbeeld het matchen van een controlegebied met een interventiegebied op relevante kenmerken. Daarnaast is het van belang om een procesevaluatie van de interventie uit te voeren waarin nauwkeurig wordt bijgehouden hoe de interventie wordt geïmplementeerd.

De Kieswijzer is geschikt voor de evaluatie van alle interventies op het grensvlak van wetenschap, praktijk en beleid in de publieke gezondheid. Complexe interventies en/of interventies met een wijk-aanpak (community interventies) nemen hierbij een speciale plaats in. Vaak



Figuur 3 Kieswijzer voor de evaluatie-opzet van interventies in de publieke gezondheid.

combineren zij meerdere soorten activiteiten in een multidisciplinaire aanpak om meerdere (tussen)doelen voor meerdere doelgroepen te bereiken. In dat geval moet de Kieswijzer meerdere malen doorlopen worden. Het is zinvol om voor elk (tussen)doel opnieuw de vraag te stellen, welke evaluatie-opzet geschikt is om te meten of dit doel wordt bereikt. Voor complexe interventies ontstaat zo een 'gelaagde' evaluatie-opzet waarbij langetermijn doelen worden gemeten in een evaluatie-opzet met controlegroep, terwijl bijvoorbeeld de samenwerking of de afzonderlijke activiteiten worden geëvalueerd in een minimale evaluatie-opzet zonder controlegroep.

VERDERE VERFIJNING WENSELIJK?

In dit opiniërend artikel hebben wij onze visie op de evaluatie-opzet gegeven en de Kieswijzer gepresenteerd. Ons inziens hebben alledrie de hoofdvarianten van de evaluatie-opzet bestaansrecht in de publieke gezondheid. De evaluatie-opzet met controlegroep is echter het meest veelbelovend, omdat zij in veel gevallen een voldoende zekere conclusie levert en bovendien flexibel kan worden aangepast aan de praktijksituatie. Het grote voordeel van deze opzet is dat men niet hoeft te randomiseren; dit biedt veel kansen om aan te sluiten bij de praktijksituatie. Ook internationaal wordt gewezen op het belang van de niet-gerandomiseerde studie-opzet voor het evalueren van interventies in de publieke gezondheid.^{2,14,23,24}

Wij hebben in dit artikel de interne validiteit van de evaluatie-opzet besproken. Hierbij zijn wij ingegaan op het uitschakelen van versturende factoren, maar niet op het uitsluiten van toeval. Beide zijn echter belangrijk om te kunnen concluderen dat de resultaten aan de interventie zijn toe te schrijven. Wanneer men toeval wil uitsluiten, moet het aantal metingen voldoende groot zijn. Het zal in de praktijk van de publieke gezondheid niet altijd haalbaar zijn om op deze manier het toeval buiten te sluiten. Dit blijft dus een factor waarmee men bij de interpretatie van de resultaten van een interventie rekening moet houden.

Naast de interne validiteit is ook de externe validiteit van de evaluatie-opzet belangrijk. De laatste hebben wij in dit artikel geheel buiten beschouwing gelaten. De externe validiteit kan worden samengevat met de vraag: kunnen wij de interventie in een andere tijd en plaats herhalen met hetzelfde resultaat? Hierbij gaat het om de vertaling van de ene setting naar de andere: van een experimentele setting naar de levensechte praktijksituatie, of van de ene maatschappelijke context naar de andere (bijvoorbeeld van de Achterhoek naar Het Gooi). Hiervoor is onder andere inzicht nodig in de mate waarin de interventie aansluit bij de bestaande cultuur en structuur en bij lokale organisaties. Het zou interessant zijn om verder in te gaan op de externe validiteit en deze ook een plaats te geven in de Kieswijzer.

Op dit moment maakt de Kieswijzer onderscheid in de drie hoofdvarianten van de evaluatie-opzet. Een verdere invulling van de Kieswijzer zou moeten ingaan op de *subvarianten* van de evaluatie-opzet. Ons inziens zijn met name de subvarianten van de evaluatie-opzet met controlegroep interessant. Het gaat hierbij dan vooral om de

keuze van de controlegroep. Men kan een externe controlegroep zoeken, maar ook een interne controlegroep of een 'historische' of 'geografische' controlegroep gebruiken.¹⁴ Wij verwachten dat de alternatieve vormen van de controlegroep goed bruikbaar zijn in de publieke gezondheid. Echter, elke subvariant heeft zijn eigen beperkingen en bijgevolg varieert de zekerheid waarmee geconcludeerd kan worden dat de resultaten aan de interventie zijn toe te schrijven. Om hierin meer inzicht te verkrijgen, zouden toekomstige evaluaties van interventies in de publieke gezondheid deze alternatieve vormen van de controlegroep moeten gebruiken.

ABSTRACT

When to use which design: 'Roadmap' for the evaluation of public health interventions

The evaluation of interventions becomes more and more important in public health. Public health practice, policy and science all benefit from an answer to the question: does the intervention work? The evaluation design that should be used to answer this question, depends on the type of intervention and the results aimed at. We are convinced that the evaluation design without control group, the evaluation design with control group and the randomised controlled trial (rct) all have the right to exist in public health. To ensure the choice of the right evaluation design, two things are of major importance. First, the outcome level one wants to measure: closeby or distant from the intervention in time and causal chain. For outcome levels closeby, the evaluation design without control group is sufficient; for outcome levels further away, a control group is necessary. The second crucial factor is how sure one wants to be about attributing the results to the intervention. The desired certainty of conclusions determines the choice between the evaluation design with control group and the rct. Based on these principles, we developed the 'Roadmap'. This 'Roadmap' facilitates the choice of the evaluation design for public health interventions.

Key words: evaluation, design, internal validity, interventions, public health, health promotion.

LITERATUUR

1. *Kuunders M, Dale D van*. Werkplan Werkplaats Certificeren 2009. Bilthoven: Centrum gezond leven, RIVM, 2008.
2. *Rychetnik L, Frommer M, Hawe P, Shiell A*. Criteria for evaluating evidence on public health interventions. *J Epidemiol Community Health* 2002;56:119-27.
3. *Anoniem*. Initiatief 'New Health Promotion'. Vier invited experts conferences met als doel: beter inzicht krijgen in hoe interventies gericht op gezondheidsbevordering het best kunnen worden geëvalueerd. Woerden: NIGZ, 2003-2004. www.newhealthpromotion.nl
4. *Alting DEM, Bouwens JGM, Keijsers JFEM*. Het evalueren van community interventies. Resultaten en conclusies van een reviewstudie naar onderzoeksmethoden en meetinstrumenten. Woerden / Den Haag: NIGZ / ZonMw, 2003.
5. *Oers H van, Verkleij H (red)*. Passend onderzoek voor gezondheidsbevordering. *Tijdschr Gezondheidswet* 2005;83:456-73.
6. *Rootman I, Goodstadt M, Hyndman B et al. (red)*. Evaluation in health promotion. Principles and perspectives. WHO, 2001.

7. *Campbell M, Fitzpatrick R, Haines A et al.* Framework for design and evaluation of complex interventions to improve health. *BMJ* 2000;321:694-6.
8. *Waters E, Doyle J, Jackson N et al.* Evaluating the effectiveness of public health interventions: the role and activities of the Cochrane Collaboration. *J Epidemiol Community Health* 2006;60:285-9.
9. *Anoniem.* The Ottawa Charter for Health Promotion. First International Conference on Health Promotion. Canada: Ottawa, 1986.
10. *Hoeijmakers M.* Typisch en volwaardig onderzoek voor gezondheidsbevordering. In: Oers H van, Verkleij H (red). *Passend onderzoek voor gezondheidsbevordering.* Tijdschr Gezondheidswet 2005;83:456-73.
11. *Wallerstein N.* A participatory evaluation model for healthier communities: Developing indicators for New Mexico. *Public Health reports* 2000;115:199-204.
12. *Last JM (editor).* A Dictionary of Epidemiology. International Epidemiological Association, Oxford, 2001.
13. *Campbell DT, Stanley JC.* Experimental and quasi-experimental designs for research. Reprinted from: *Handbook of Research on Teaching.* Houghton Mifflin Company, 1963.
14. *Habicht JP, Victora CG, Vaughan JP.* Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact. *Int J Epidemiol* 1999;28:10-8.
15. *Hawe P, Shiell A, Riley T.* Complex interventions: how "out of control" can a randomised controlled trial be? *BMJ* 2004;328:1561-3.
16. *Jackson N.* Systematic reviews of health promotion and public health interventions: Handbook. *Cochrane Health Promotion and Public Health Field*, 2005.
17. *Yperen TA van, Veerman JW (red).* Zicht op effectiviteit. *Handboek voor praktijkgestuurd effectonderzoek in de jeugdzorg.* Delft: Eburon, 2008.
18. *Koelen MA, Vaandrager L, Colomer C.* Health promotion research: dilemma's and challenges. *J Epidemiol Community Health* 2001;55:257-62.
19. *Nutbeam D.* Evaluating health promotion – progress, problems and solutions. *Health Promotion Int* 1998;13:27-44.
20. *Saan H, Haes W de.* *Gezond effect bevorderen. Het organiseren van effectieve gezondheidsbevordering.* Woerden: NIGZ, 2005.
21. *Anoniem.* GGD Masterclass 'Samenwerking over de grenzen heen' groot succes. Utrecht: GGD Nederland, 2008. www.ggd Kennisnet.nl
22. *Jansen SC, Lezwijn J, Schreurs H, Rots C.* Themamiddag: Epidemiologie en gezondheidsbevordering op weg naar synergie. Utrecht: GGD Nederland, 2007. www.ggd Kennisnet.nl
23. *Victoria CG, Habicht JP, Bryce J.* Evidence-based public health: moving beyond randomized trials. *Am J Public Health* 2004;94:400-5.
24. *Des Jarlais DC, Lyles C, Crepaz N et al.* Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. *Am J Public Health* 2004;94:361-6.
25. *Izeboud C, Rover C de, Vlaming R de, Haveman-Nies A.* Het project 'Alcoholmatiging Jeugd in de Achterhoek': de ontwikkeling van een complexe interventie. *Tijdschr Gezondheidswet* 2008;86:405-11.
26. *Vlaming R de, Rover C de, Izeboud C, Jansen SC, Wit I de, Haveman-Nies A.* Evaluatie van 'Alcoholmatiging Jeugd in de Achterhoek': verandering in perceptie van ouders op het alcoholgebruik van jongeren na één jaar. *Tijdschr Gezondheidswet* 2008;86:412-21.

CORRESPONDENTIEADRES

**S.C. van Oord-Jansen,
GGD Gelre-IJssel, afdeling KEC,
Postbus 51, 7300 AB Apeldoorn,
tel. 088-4433218,
e-mail: s.jansen@ggdgelre-ijssel.nl.**