

J. A. KNOTTNERUS EN A. VOLOVICS

Over validiteit, vergelijkbaarheid en versturende variabelen

Een huisarts meldt zich bij een universitair huisartseninstituut met de vraag of men hem wil helpen met de statistische bewerking van zijn materiaal: hij blijkt al vijf jaar te hebben bijgehouden wie volgens een door hem geformuleerd hypertensieprotocol wordt behandeld, en hoe het beloop in die gevallen is. Hij wil nu de effectiviteit van zijn aanpak vaststellen. Eventueel zijn, met enige extra inspanning, bloeddrukgegevens te verkrijgen van een 'controlegroep' van hypertensiepatiënten uit enkele naburige praktijken bij wie géén speciaal protocol is toegepast. Tot zijn teleurstelling krijgt de huisarts te horen dat zijn onderzoeksvraag op grond van de verzamelde gegevens niet goed te beantwoorden is. Zou hij zijn tijd niet beter kunnen besteden aan het opzetten van een nieuw onderzoek, zo wordt geopperd?

Inleiding

Eerder in deze serie kwam naar voren dat statistische analyse en toetsing van onderzoeksgegevens alleen zin heeft, als die gegevens het resultaat zijn van een 'valide' studie: een onderzoek dat zodanig is opgezet en uitgevoerd, dat het antwoord op de onderzoeksvraag niet wordt vertroebeld door een of andere vorm van vertekening ('bias'). In deze aflevering zullen we enkele belangrijke vormen van bias bespreken. Enig inzicht hierin is onontbeerlijk, zowel bij het zelf opzetten van onderzoek en het analyseren van de uitkomsten, als bij het kritisch lezen van onderzoekspublikaties.

Een vruchtbaar vertrekpunt hierbij is het besef dat huisartsgeneeskundig wetenschappelijk onderzoek, bijvoorbeeld op het gebied van etiologie, diagnostiek, therapie en prognose, voor een belangrijk deel bestaat uit het vergelijken van verschillende groepen of van verschillende toestanden van dezelfde groep bij verschillende gelegenheden. Bij dit vergelijken dient dan te zijn voldaan aan de voorwaarde van *a priori* vergelijkbaarheid: afgezien van de te bestuderen factoren mogen er vooraf geen relevante verschillen zijn tussen de

te vergelijken groepen wat betreft het voorkomen van andere factoren die de te onderzoeken uitkomst kunnen beïnvloeden. Op zijn minst moet in de analyse van de resultaten gecorrigeerd kunnen worden voor dergelijke verschillen.

Wij besteden aandacht aan de vertekening die kan ontstaan door tekortkomingen met betrekking tot:

- de vergelijkbaarheid van de vergeleken groepen;
- de vergelijkbaarheid van de vergeleken effecten;
- de vergelijkbaarheid van informatieverwerving.

Vervolgens worden nog enkele, in dit verband dikwijls gehanteerde, begrippen nader besproken: versturende variabele ('confounder'), selectiebias en effectmodificatie.

Het belang van vergelijken

Voor het kunnen trekken van conclusies uit een onderzoek is het meestal nodig dat er een vergelijking is ingebouwd. *Van der Wal* wijst in dit verband op een beperking van zijn onderzoek naar risicofactoren voor het krijgen van *pelvic inflammatory disease* (PID). De meeste PID-patiënten zeiden dat zij in de maanden voorafgaande aan de PID geconfronteerd waren geweest met een grote lichamelijke en vooral emotionele belasting (examens, dubbele banen en studies, onregelmatige diensten).¹ Voor dat dit tot risicofactor voor PID wordt verklaard, moet echter eerst worden vastgesteld of een dergelijke belasting zich duidelijk minder vaak heeft voorgedaan bij IUD-gebruiksters die géén PID hebben ontwikkeld; het is immers denkbaar dat beide groepen jonge vrouwen het in het algemeen zwaar hebben, los van het eventueel krijgen van PID.

Ook de huisarts in de inleidende casus zal het niet zonder vergelijkings- of 'controle'-groep kunnen stellen. Om een zinvolle uitspraak te kunnen doen over de effectiviteit van zijn hypertensieprotocol, kan hij niet volstaan met kennis van het verloop van de bloeddruk en/of het klinisch beloop bij de patiënten die volgens het protocol wor-

den behandeld. Ook kennis van het verloop bij niet of 'anders' behandelde hypertensiepatiënten is nodig, al was het alleen al omdat ook personen die meestal een lagere bloeddruk hebben, tot de hoge-bloeddrukpatiënten kunnen worden gerekend. Ten gevolge van de biologische variatie van de bloeddruk kunnen zij op het moment van de meting een toevallige uitschieter boven de gehanteerde grenswaarde hebben gehad. De kans is groot dat bij deze mensen later - onafhankelijk van de behandeling - een lagere bloeddruk wordt gevonden, meer in de buurt van het persoonlijk gemiddelde (regressie naar het gemiddelde). Vergelijking met een geschikte controlegroep van hoge-bloeddrukpatiënten, waarin dit verschijnsel zich uiteraard ook voordoet, maakt het mogelijk het hierdoor optredende schijneffect van de behandeling 'weg te strepen'. Dat kan dan bovendien gebeuren met betrekking tot de invloed van allerlei andere factoren die niets te maken hebben met de specifieke kenmerken van het protocol, zoals de wetenschap hoge bloeddruk te hebben en de aandacht van de arts op zichzelf.

Vergelijkbaarheid van groepen

Als men eenmaal een controlegroep heeft, is men er nog niet. Zoals gezegd, moeten eisen worden gesteld ten aanzien van de *a priori* vergelijkbaarheid van beide groepen. Als het gezochte verband of effect niet bestaat, zouden de uitkomsten van beide groepen, eventueel na correctie voor andere factoren, niet mogen verschillen.

Het belang hiervan kan worden geïllustreerd worden aan de hand van twee vaak gestelde vragen:

- Leidt praktijkverkleining tot minder verwijzen door de huisarts?^{2 3}
- Leidt het werken in een gezondheidscentrum tot minder verwijzen?

Onderzoek op dit gebied bestaat in de regel uit het vergelijken van het verwijsgedrag van twee groepen huisartsen: enerzijds huisartsen met kleinere of verkleinde praktijken of huisartsen in gezondheidscentra, en anderzijds huis-

Rijksuniversiteit Limburg, Postbus 616, 6200 MD Maastricht.

Prof. dr. J.A. Knottnerus, hoogleraar huisartsgeneeskunde, vakgroep Huisartsgeneeskunde; A. Volovics, statisticus, vakgroep Medische Informatica en Statistiek.

Correspondentie: Prof. dr. J.A. Knottnerus.

artsen in grotere praktijken of solisten. Het probleem hierbij is dat er veel meer voor het verwijsgedrag relevante verschillen zijn tussen de vergeleken groepen. Voor een aantal van die verschillen is eigenlijk niet bevredigend te corrigeren, zo ze al te meten zijn. Een belangrijk aspect is bijvoorbeeld dat in kleine praktijken en in gezondheidscentra veelal artsen met een andere attitude (gaan) werken dan in andere praktijken. Voorts is het mogelijk dat dergelijke praktijken een ander type patiënten aantrekken.

Het vraagstuk van de vergelijkbaarheid van groepen doet zich ook voor bij de meting van de effecten van nascholing en toetsing. Meestal is men dan aangewezen op het vergelijken van de vorderingen van huisartsen die een nascholings- of toetsingsprogramma hebben gevolgd met de resultaten van collega's die daaraan niet hebben deelgenomen.⁴ Het is in zulke gevallen zeer wel denkbaar dat er vooraf (moeilijk meetbare) verschillen in attitude bestaan tussen beide groepen. Een eventueel gevonden verschil in vorderingen kan hiermee samenhangen.

Bij patiëntgebonden effectiviteitsonderzoek komt de vergelijkbaarheid in het gedrag als de indeling in groepen wordt bepaald door de indicatiestelling van de arts. Deze wordt immers sterk beïnvloed door de ernst van de aandoening, contra-indicaties en de vermeende effectiviteit. In een recent retrospectief onderzoek naar de invloed van bètablokkers op de kwaliteit van het leven ging men uit van patiënten bij wie door de huisarts in het verleden een behandeling met een bètablokker of diureticum was ingesteld.⁵ In zo'n geval moet men bedacht zijn op zogenaamde *confounding by indication*. Het ligt bijvoorbeeld voor de hand dat bij ernstiger of meer riskante vormen van hypertensie vaker middelen (of combinaties van middelen) zijn voorgeschreven waarvan een sterker effect werd verwacht. Ook is denkbaar dat bij personen met een minder goede cardiopulmonale status of met veel bijwerkingen in het verleden over het algemeen andere middelen werden voorgeschreven dan bij personen zonder dergelijke klachten. Bij het vergelijken van de effectiviteit en de veiligheid van door de arts voorgeschreven middelen hebben bepaalde middelen dan een ongunstiger startpositie en hun werkzaamheid kan bij dergelijk on-

derzoek worden gemaskeerd (tabel 1).

Confounding by indication kan ook effectiviteit suggereren of te sterk doen lijken. Een goed voorbeeld is de situatie waarin het besluit om al dan niet te verwijzen voor operatieve therapie in sterke mate wordt bepaald door contra-indicaties (bijvoorbeeld een slechte algemene conditie). Zo zullen voor een operatie van een lokaal vaatlijden alleen patiënten worden verwezen die niet tevens een uitgebreid algemeen vaatlijden hebben. Bij vergelijking van het beloop

van geopereerden en niet-geopereerden hebben de eersten dan een betere uitgangssituatie. Hier is sprake van een variant op het *healthy worker effect*: veel risico's van mogelijk gezondheidsbedreigende werksituaties zijn moeilijk of niet aan te tonen, doordat alleen de gezondste werknemers tot deze werksituaties worden toegelaten (via de aanstellingskeuring). Ook voor deze vormen van vertekening is in de praktijk dikwijls niet te corrigeren, doordat de factoren die de indicatiestelling hebben

Tabel 1 *Confounding by indication: effectiviteit en veiligheid van twee geneesmiddelen bij hypertensie.*

Type hypertensie	N	Normotensief		Bloeddrukdaling na 6 maanden mm Hg	Bijwerkingen	
		N	%		N	%
<i>Middel A</i>						
Resistent	90	36	40	14	14	16
Niet-resistent	10	9	90	20	1	10
Totaal	100	45	45	15	15	15
<i>Middel B</i>						
Resistent	10	2	20	12	2	20
Niet-resistent	90	54	60	17	12	13
Totaal	100	56	56	17	14	14

Zowel bij 'resistente' als minder resistente vormen van hypertensie is middel A in werkelijkheid effectiever en veiliger dan middel B. Omdat middel A echter veel vaker bij resistente vormen wordt voorgeschreven, lijken in totaal de resultaten bij personen die middel B kregen gunstiger.

Tabel 2 *Voorbeelden van verschillende effect-vergelijkingen (contrasten).*

A Bij onderzoek naar de effectiviteit van een protocol X		
- protocol X	↔	geen protocol
- protocol X	↔	een ander protocol
- protocol X	↔	diagnostiek en begeleiding als tot nu toe gebruikelijk
B Bij onderzoek naar schadelijke effecten van vogelhouden		
- vogelhouden	↔	niet vogelhouden
- aanwezigheid van zieke vogels	↔	geen zieke vogels
- goede kooihygiëne	↔	geen goede kooihygiëne
- stofbelasting	↔	geen stofbelasting
C Bij onderzoek naar de effectiviteit van hartbewaking bij (mogelijk) myocardinfarct		
- opname op HBE	↔	opname elders
- opname op HBE	↔	thuis blijven
- geavanceerde diagnostiek	↔	niet-geavanceerde diagnostiek
- geavanceerde therapie	↔	niet-geavanceerde therapie
- gebruik bewakingsapparatuur	↔	geen gebruik bewakingsapparatuur
- intensieve verpleging	↔	niet-intensieve verpleging
- gespecialiseerd personeel	↔	niet-gespecialiseerd personeel

bepaald, niet gemeten of onvoldoende bekend zijn.

Apart te noemen is nog het probleem van de uitval van patiënten tijdens een onderzoek, bijvoorbeeld tijdens een onderzoek naar de effectiviteit van een geneesmiddel; men vergelijkt hierbij het beloop in groep A, die de kuur met het middel heeft afgemaakt, met het beloop van groep B, die in dezelfde periode het middel niet kreeg. Als nu de patiënten die de kuur *niet* hebben afgemaakt vanwege bijwerkingen of omdat het middel niet hielp – de ongunstige gevallen dus – niet in groep A worden meegeteld, geeft dit een vertekening ten gunste van groep A, als een dergelijke uitval niet of minder heeft plaatsgevonden in groep B. Het is niet uitgesloten dat deze *differentiële uitval* een rol heeft gespeeld in het aangehaalde onderzoek over bètablokkers, zoals de auteurs ook stellen.⁵

Een probleem dat regelmatig opduikt, is het conflict tussen het vaststellen van de werkzaamheid van een middel bij maximale participatie (dus zonder uitval) en het nagaan van de effectiviteit in de praktijk (waarbij ook eventuele uitval als maat voor non-acceptatie van het middel moet worden ingecalculleerd). Omdat men in het eerste geval juist speciale maatregelen zal nemen om de uitval te voorkomen, zijn beide onderzoeksdoelen niet via eenzelfde onderzoek(sopzet) te bereiken.

Vergelijkbaarheid van effecten

Bij het bestuderen van het effect van een etiologische of therapeutische factor verschaft alleen een nauwkeurig omschreven vraagstelling de sleutel tot een valide onderzoeksopzet. Waarmee wil men het effect vergelijken? Welk *contrast* streeft men na? Wil de huisarts in de inleidende casus zijn hypertensieprotocol vergelijken met het werken zonder een protocol, of met de aanpak zoals huisartsen die tot nu toe gewoon zijn? Of wil bepalen welke meerwaarde zijn protocol heeft ten opzichte van een ander protocol?

Het is duidelijk dat elk van deze gevallen een specifieke vergelijking nodig maakt, met telkens een andere 'controle-interventie' (tabel 2a). Als men conclusies trekt over een bepaalde vergelijking, terwijl in werkelijkheid een andere vergelijking is onderzocht, spreekt men 'onvergelijkbaarheid van effecten'.

Het is ook van belang om vooraf vast te stellen in hoeverre gecorrigeerd moet worden voor het aspect 'aandacht van de arts voor de patiënt'. Dit gaat bijvoorbeeld een rol spelen, als de patiënt in het kader van een CARA-protocol om de drie maanden systematisch moet worden onderzocht, terwijl dit anders niet of minder systematisch zou gebeuren. Dit corrigeren is niet alleen een praktisch-methodologisch probleem (kun je een soort placebo-protocol realiseren?), het is ook een principiële kwestie: wat zijn de kenmerken van het protocol ten opzichte van andere handelwijzen; is het aspect 'extra aandacht' hiervan een wezenlijk element of wordt alleen een specifiek medisch-technische meerwaarde gepretendeerd?

Dergelijke vragen kan men zich ook stellen bij onderzoek van etiologische factoren als de schadelijkheid van huisvogels,⁶ of de effectiviteit van voorzieningen als hartbewakingseenheden (tabel 2b-c). Dit is niet alleen belangrijk voor het beantwoorden van de vraagstelling maar ook voor de praktijk: als het lukt om de effectiviteit van een voorziening als de hartbewakingseenheid te ontleden in haar effectieve elementen, kan beter bepaald worden welke nuttige onderdelen van deze voorziening ook elders (bijvoorbeeld thuis) zijn te realiseren.

Vergelijkbaarheid van informatieverwerving

Uitkomsten van een vergelijkend onderzoek kunnen ook vertekend zijn door toepassing van verschillende methoden van waarneming in de vergeleken groepen. In de inleidende casus zijn de bloeddruk en andere klinische gegevens van de volgens protocol behandelde patiënten prospectief en systematisch bijgehouden; er is gemeten met één bloeddrukmeter, en er was één enthousiaste huisarts die de bloeddrukwaarden aflas en noteerde. Als later een controlegroep uit naburige praktijken zou worden samengesteld, zouden de gegevens achteraf daarvoor moeten worden teruggezocht op patiëntenkaarten; bovendien zijn er andere bloeddrukmeters gebruikt door andere, wellicht minder gemotiveerde huisartsen. Het ligt dus voor de hand dat er tussen beide groepen verschillen zijn in de betrouwbaarheid van de metingen en de volledigheid van de patiëntengegevens,

en dat hierdoor vertekening optreedt (*informatiebias*). Het risico is groot dat de gegevens van de probleemloos op therapie reagerende patiënten relatief vaak zullen ontbreken in de controlegroep, en dat deze er dus te ongunstig afkomt.

De vertekening zit hem niet alleen in de beperkte kwaliteit van de gehanteerde meetmethoden, maar vooral in het *verschil* in meetmethoden tussen de twee groepen. Beide hoeven niet samen te gaan. In het kader van een onderzoek in een huisartspraktijk naar factoren die invloed hebben op de digoxinespiegel werden achteraf – na de bepaling van het verloop van de digoxinespiegels – alle patiënten ondervraagd over de hoeveelheid dagelijkse lichaamsbeweging tijdens het onderzoek.⁷ Hoewel de retrospectieve aanpak de kans op 'meetfouten' verhoogt, is dit verschijnsel in beide groepen in gelijke mate te verwachten. Hierbij is ook van belang dat de patiënten bij het beantwoorden van deze vraag geen kennis hadden van hun inmiddels al gemeten digoxinewaarden; zou dit namelijk in de ene groep vaker het geval zijn dan in de andere, dan zou er informatiebias kunnen optreden: de ene groep zou in zijn antwoord meer beïnvloed kunnen zijn door de voorkennis dan de andere. Het ligt in dit onderzoek al met al niet voor de hand dat het gevonden negatieve verband tussen lichaamsbeweging en digoxinespiegel het gevolg is van informatiebias. Overigens geldt in het algemeen natuurlijk dat een meetmethode met een minimale foutenkans de voorkeur verdient, omdat een eventueel bestaand verband dan gemakkelijker en nauwkeuriger wordt aangetoond.

Het risico op onvergelijkbaarheid van informatieverwerving ligt vaak op de loer bij onderzoek naar de samenhang tussen diagnostische gegevens en bepaalde aandoeningen. Als men geen systematisch diagnostisch onderzoeksprotocol volgt, is de kans groot dat bij aanwezigheid van symptomen intensiever naar de aandoening wordt gezocht. Alleen al hierdoor kan de ziekte in deze groep vaker worden aangetroffen.⁸ Men spreekt dan wel van *detectiebias*.

Soms is het aannemelijk dat het eindresultaat wordt beïnvloed door de waarnemingsmethode zelf. Dat zal zich bijvoorbeeld voordoen bij een onderzoek naar de effectiviteit van een bepaald

gewichtsverminderend dieet, waarbij de patiënten om de twee weken worden gewogen door de doktersassistente. Bij de controlegroep (die het dieet niet krijgt) zal men dat dan ook moeten als men het specifieke effect van het dieet wil onderscheiden van de invloed van het regelmatig gewogen worden. Behalve de kwestie van het placebo-effect van het dieet vraagt dus ook het probleem van het *placebo-effect van de meetmethode* om een oplossing.

Meer over verstorende variabelen

Factoren die onvergelykbaar zijn op één van de hierboven beschreven wijzen, worden ‘verstorende variabelen’ of ‘confounders’ genoemd. Zij hebben twee kenmerken:

- zij hangen samen met de onafhankelijke variabele X;
- zij beïnvloeden de afhankelijke variabele Y (de uitkomst) (figuur).

Een verstorende variabele kan een aanwezig verband maskeren, of een zwak of zelfs afwezig verband versterken c.q. suggereren, zoals blijkt uit de volgende voorbeelden.

Rutten e.a. bespraken een groot aantal onderzoeken naar de effectiviteit van verschillende behandelingen van lage rugpijn.⁹ Het bleek dat meestal geen rekening was gehouden met het feit dat patiënten vaak meer dan één therapie tegelijk krijgen. Als een behandeling waarop het onderzoek *niet* primair is gericht, ongelijk is verdeeld over de vergeleken groepen, is er echter sprake van ‘onvergelykbaarheid van effecten’. Het is dan zelfs denkbaar dat men een effect vindt, dat in feite geheel berust op de effectiviteit van de andere therapie (de *verstorende variabele*) (tabel 3).

Een voorbeeld van potentiële confounding op het gebied van etiologisch onderzoek is het verband tussen obesitas en sterfterisico met als verstorende variabele het roken.¹⁰ Rokers zijn over het algemeen lichter dan niet-rokers, dus onder mensen met obesitas komen relatief minder rokers voor. Als men hiermee geen rekening houdt in de analyse, is er sprake van onvergelykbaarheid van groepen, en kan daardoor het verband tussen obesitas en sterfte worden gemaskeerd: men vindt een te laag *overall* relatief risico, doordat een deel van het verhoogde risico bij overgewicht wordt gecompenseerd door minder roken. Als men het relatief risico apart

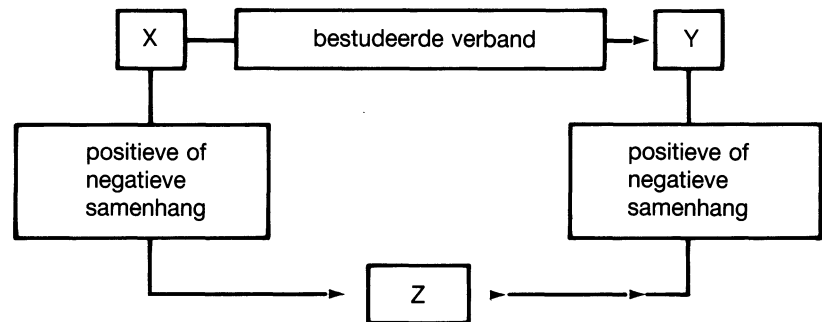
berekent voor rokers en niet-rokers, dan blijkt dit telkens hoger te zijn (tabel 4). Dit apart berekenen van het relatief risico van obesitas voor rokers en niet-rokers is een kenmerk van een *gestratificeerde analyse*, waarmee dus ‘gecorrigeerd’ wordt voor een verstorende variabele.

Men moet zich, alvorens tot een dergelijke correctie over te gaan, steeds afvragen of de betreffende factor wel echt een verstorende variabele is. De factor kan namelijk een onderdeel zijn van het mechanisme waarlangs de primair bestudeerde etiologische factor werkt. Zo verloopt een deel van de nadelige werking van overgewicht waarschijnlijk via een verhoogde bloeddruk.¹⁰ Iets dergelijks geldt voor de ongezonde invloed van stress die bij veel mensen verloopt via het roken van

(meer) sigaretten. Als men voor een dergelijke vermeende ‘verstorende variabele’ gaat corrigeren, kan dit leiden tot een onderschatting van het onderzochte risico (tabel 5). Er is dan sprake van *overcorrectie*.

Selectiebias is een vorm van onvergelykbaarheid van groepen en de verstorende variabele is dan de factor die verantwoordelijk is voor de selectiebias. Van der Boog e.a. vonden dat kinderen die de polikliniek voor groeistoornissen bezochten, relatief vaak particulier verzekerd waren: 57 procent versus 32 procent van de kinderen die in dezelfde tijd de algemene polikliniek voor kinderziekten bezochten.¹¹ Men zou nu kunnen concluderen dat er een verband is tussen het voorkomen van groeiproblemen en verzekeringsvorm. In dat geval is men waarschijnlijk het slachtoffer van

Figuur De relatie tussen enerzijds de verstorende variabele Z en anderzijds de onafhankelijke variabele X en de afhankelijke variabele Y.



Tabel 3 Vergelijking van twee groepen van 100 patiënten, wel en niet behandeld met behandeling I.

A Overall 'effectiviteit' van behandeling I

Behandeling I	Klachten over		Totaal aantal
	ja	nee	
+	72	28	100
-	52	48	100

Resultaat in de groep met behandeling I: bij 72 procent gunstig. Resultaat in de groep zonder behandeling I: bij 52 procent gunstig. Behandeling I is schijnbaar effectief, doch dit effect berust geheel op een tweede, niet-bestudeerde behandeling II, die vaker wordt toegepast bij patiënten die ook behandeling I krijgen. Behandeling II is een ‘verstorende variabele’.

B Onderzoeksresultaat van dezelfde studie indien het effect van behandeling II in de analyse zou zijn betrokken

Behandeling		Klachten over		Totaal aantal
I	II	ja	nee	
+	+	64	16	80
+	-	8	12	20
-	+	24	6	30
-	-	28	42	70

Resultaat in de groepen met behandeling II, met of zonder behandeling I: bij 80 procent gunstig. Resultaat in de groepen zonder behandeling II, met of zonder behandeling I: bij 40 procent gunstig.

Tabel 4 Verband tussen obesitas en sterftecijfer, met als versturende variabele roken (voorbeeld).

A Verband tussen obesitas en sterfte onder 10.000 personen, zonder rekening te houden met roken

	Overleden	Niet overleden	Totaal
Obesitas	30	2.970	3.000
Geen obesitas	50	6.950	7.000
Totaal	80	9.920	10.000

$$\text{Relatief risico} = \frac{30/3.000}{50/7.000} = 1,4$$

B Verband tussen obesitas en sterfte per jaar in dezelfde populatie, apart voor rokers en niet rokers

	Rokers			Niet rokers		
	Overleden	Niet overleden	Totaal	Overleden	Niet overleden	Totaal
Obesitas	17	733	750	13	2.237	2.250
Geen obesitas	40	3.460	3.500	10	3.490	3.500
Totaal	57	4.193	4.250	23	5.727	5.750

$$\text{Relatief risico bij rokers} = \frac{17/750}{40/3.500} = 2,0$$

$$\text{Relatief risico bij niet-rokers} = \frac{13/2.250}{10/3.500} = 2,0$$

Als men geen rekening houdt met roken, valt het relatief risico te laag uit. Onder niet-obesen wordt meer gerookt, en roken verhoogt het sterfterisico.

Tabel 5 Overcorrectie voor een vermeende versturende variabele (voorbeeld).

A Overall-verband

	Myocardinfarct (per jaar)		
	+	-	Totaal
Stress	25	3.975	4.000
Geen stress	15	5.985	6.000
Totaal	40	9.960	10.000

$$\text{Relatief risico} = \frac{25/4.000}{15/6.000} = 2,5$$

B Apart voor rokers en niet-rokers

	Rokers			Niet rokers		
	Infarct	Geen infarct	Totaal	Infarct	Geen infarct	Totaal
Stress	23	2.977	3.000	2	998	1.000
Geen stress	10	1.990	2.000	5	3.995	4.000
Totaal	33	4.967	5.000	7	4.993	5.000

$$\text{Relatief risico bij rokers} = \frac{23/3.000}{10/2.000} = 1,5$$

$$\text{Relatief risico bij niet-rokers} = \frac{2/1.000}{5/4.000} = 1,6$$

selectiebias, waarbij de frequentie van het bezoeken van de polikliniek voor groeistoornissen de boosdoener is (tabel 6). De onderzoekers zoeken de verklaring namelijk in een sterkere neiging van particuliere verzekerden om het voor dit probleem 'hogerop' te zoeken. Deze hypothese wordt gesteund door het feit dat particulier verzekerden relatief vaak geen verwijfsbrief van de huisarts bij zich hadden.

Meestal vraagt het corrigeren van selectievertekening meer dan het toepassen van een statistische procedure. Behalve de selecterende factor moet ook de mate waarin selectie naar de polikliniek is opgetreden, bekend zijn. In het algemeen ontbreekt informatie hierover en kan alleen een andere onderzoekspopzet redding brengen.⁸

Selectieproblemen liggen op de loer bij het doen van patiënt-controle onderzoek op basis van huisartsgeneeskundige registratiesystemen. Als men geregistreerde patiënten met een bepaalde aandoening als onderzoeksgroep definieert, gaat men uit van via het dagelijks spreekuur en niet via standaard-procedures gediagnostiseerde gevallen. Deze gevallen kunnen sterk geselecteerd zijn ten opzichte van de niet-vastgestelde gevallen, waarbij immers allerlei factoren *niet* zijn onderzocht. Het is dan niet eenvoudig een geschikte controlegroep te vinden. Dit probleem is vooral aan de orde bij aandoeningen die vaak onder de oppervlakte blijven, zoals sinusitis of migraine. Bij ziekten die vrijwel altijd zonder veel omwegen in het register komen, zijn meestal goede oplossingen te vinden.¹² In deze gevallen is het al dan niet in de register terecht komen de versturende variabele.

Het is ook mogelijk dat de patiënten die in een onderzoeksgroep worden opgenomen, allen gestandaardiseerde diagnostiek ondergaan, terwijl dit bij de uit het register getrokken controlepatiënten niet is gebeurd.¹³ Men moet dan bedacht zijn op 'onvergelijkbaarheid van informatieverwerking'. Versturende variabele is dan de mate waarin de onderzoeksgroepen op de *bestudeerde factoren zijn gescreend*.

Omdat registratiesystemen bij voortschrijdende computerisering alleen maar belangrijker zullen worden als informatiebron, behoren de methodologische vraagstukken op dit terrein tot de meest dringende uitdagingen voor de huisartsgeneeskundige wetenschap.

Effect-modificatie

Het is belangrijk om onderscheid te maken tussen versturende en *modificerende* variabelen. De laatste zijn persoonsgebonden kenmerken, waardoor een bepaald verband 'gemodificeerd' wordt, dat wil zeggen: sterker of minder sterk uitvalt.

Vaak is het geslacht een modificerende variabele, bijvoorbeeld bij het verband tussen de leeftijd en de incidentie van borstkanker. Onder vrouwen is dit verband immers veel sterker: de incidentie neemt duidelijk toe met de leeftijd. Bij mannen is de incidentie ook op hoge leeftijd zeer laag. De leeftijd zelf is ook vaak een modificerende variabele. Een duidelijk voorbeeld van modificatie door de leeftijd zien we bij het verband tussen geslacht en risico op ischemische hartziekten: op hoge leeftijd halen de vrouwen de mannen in.^{14 15}

Er is hier geen sprake van verstoring van onderzoeksresultaten; veeleer geeft het begrip modificatie een sleutel tot het benaderen van de rijk geschakeerde realiteit en meer individualisering van de uitkomsten. Overigens kan een modificerende variabele die ongelijk verdeeld is over de vergelijkingsgroepen, wél als versturende variabele werken: als in een onderzoek naar het verband tussen leeftijd en incidentie van borstkanker onder de ouderen relatief meer mannen voorkomen dan onder de jongeren, dan zou het bestudeerde verband gemaskeerd kunnen worden; men moet daar in het onderzoek dus rekening mee houden.

Slotopmerkingen

Er zijn allerlei statistische technieken om tegelijkertijd te corrigeren voor de invloed van diverse versturende variabelen en tevens de invloed van modificerende variabelen na te gaan. Alle hebben ze hun premissen en voor- en nadelen. In een volgende aflevering komen we op de principes van deze technieken terug. Twee opmerkingen zijn hier nog van belang:

- Correctie, volgens welke techniek ook, is alleen mogelijk voor variabelen die zijn gemeten en dus tevoren reeds zijn geïdentificeerd.
- Versturende variabelen waarop men niet heeft gerekend en die niet zijn gemeten, kunnen dus niet in de analyse worden betrokken.

Over de consequenties van dit laatste

Tabel 6 Selectiebias: verband tussen verzekeringsvorm en groeistoornissen (voorbeeld).

A Kinderen in de algemene bevolking

	Groeiproblemen		Geen groeiproblemen	
	N	%	N	%
Particulier	316	31,6	3.160	31,6
Ziekenfonds	684	68,4	6.840	68,4
Totaal	1.000	100	10.000	100

B Percentage dat polikliniek bezocht per subgroep

	Groeiproblemen	Geen groeiproblemen
Particulier	36,0	12,4
Ziekenfonds	12,4	12,4

C Kinderen op de polikliniek kinderziekten

	Groeiproblemen		Geen groeiproblemen	
	N	%	N	%
Particulier	114	57,3	392	31,6
Ziekenfonds	85	42,7	848	68,4
Totaal	199	100	1.240	100

Doordat particulier verzekerde kinderen veel vaker de polikliniek (voor groeistoornissen) bezoeken, ontstaat er op de polikliniek kinderziekten een schijnverband tussen particulier verzekerd zijn en groeiproblemen.

blijft, met name als de onderzoeksgroepen niet door loting zijn samengesteld, onzekerheid bestaan. Omdat randomisatie lang niet altijd mogelijk is, en ook anderszins de 'ideale' onderzoeksopzet vaak hoogstens benaderd kan worden, is deze onzekerheid binnen bepaalde grenzen acceptabel. In elk geval mag verwacht worden dat de onderzoeker motiveert waarom hij bepaalde keuzen – met inachtneming van hun beperkingen – heeft gemaakt, en dat hij bespreekt welke consequenties dit heeft voor de interpretatie van de resultaten.

¹ Van der Wal G. IUD en pelvic inflammatory disease. *Huisarts Wet* 1985; 28: 11-5.
² Van der Meulen P, Spreuwenberg C. Praktijkverkleining en de tijdsbesteding van de huisarts. Een literatuurstudie. *Huisarts Wet* 1987; 30: 76-8.
³ Post D, Gubbels JW. Verwijzen naar interne specialismen. Verschillen tussen huisartsen en enkele factoren die daarop van invloed zijn. *Huisarts Wet* 1986; 29: 369-72.
⁴ Grol R, Mokkink H, Schellevis F, Van Deursen J, Teunissen A. Effecten van onderlinge toetsing. *Huisarts Wet* 1986; 29: 99-103; 141-5.
⁵ Tempelaar AF, Haayer-Ruskamp FM, Pennink BJ, Smith RAJ, Waaijer AMM, Wesseling H. Langdurig gebruik van bèta-

blokkers en de kwaliteit van het leven. Wordt gepubliceerd in mei of juni van de lopende jaargang van *Huisarts en Wetenschap*.

⁶ Holst PAJ. Gezondheidsrisico's van huisvogels [Dissertatie]. Utrecht: Rijksuniversiteit Utrecht, 1987.
⁷ Meyboom WA. Digoxine in een huisartspraktijk. *Huisarts Wet* 1987; 30: 303-7.
⁸ Knottnerus JA, Knipschild PG, Sturmans F. Symptomatie en selectiebias. *Huisarts Wet* 1985; 28:325-330.
⁹ Rutten G, Faas A, Chavannes A. De behandeling van lage rugpijn: wat is zin-nig? *Huisarts Wet* 1987; 30: 346-8.
¹⁰ Seidell JC, Deurenberg P, Hautvast JGAJ. Te dik, en wat dan nog. *Huisarts Wet* 1987; 30(suppl 11): 42-5.
¹¹ Van der Boog SJ, Van der Spek RN, Steendijk R. Verwijzingen naar een polikliniek voor groeistoornissen. *Huisarts Wet* 1987; 30: 141.
¹² Knottnerus JA. Subject selection in hospital based case-control studies. *J Chron Dis* 1987; 40: 183-5.
¹³ Lagro-Janssen T. Nooit op tijd. Een onderzoek naar urine-incontinentie bij de vrouw. *Huisarts Wet* 1986; 29: 38-41.
¹⁴ Lamberts H, Van der Velden HGM. Ziekten van hart en vaten in de huisartspraktijk. Van naïef positivisme naar kritisch realisme. *Huisarts Wet* 1987; 30(suppl 11): 4-14.
¹⁵ IJzermans CJ. Hart, vaten en gezin. *Huisarts Wet* 1987; 30(suppl 11): 31-6.