

Waarnemingsmethoden in kwalitatief onderzoek

Tony Hak

Inleiding

Bij de term kwalitatief onderzoek denkt men vrijwel meteen aan bepaalde vormen van gegevensverzameling of 'waarneming', de term die ik in deze bijdrage zal gebruiken, zoals diepte-interviews en focusgroepen. Toen de BMJ in 1995 voor het eerst een serie artikelen wijdde aan kwalitatief onderzoek – zoals H&W nu – werden er dan ook aparte bijdrages gewijd aan het kwalitatieve interview en aan de focusgroep.^{1,2} Dit waren lange tijd de enige twee vormen van onderzoek die erin slaagden het peer-reviewproces succesvol te doorlopen. Hoewel ook observatie,³ consensusmethodes⁴ en de gevalsbeschrijving⁵ in deze artikelenserie aan bod waren gekomen, kwamen deze vormen van onderzoek niet in het BMJ voor tot eind 2000, toen er voor het eerst een observationeel onderzoek werd gepubliceerd.^{6,7}

De BMJ is een voortrekker, maar ook bij andere medische tijdschriften die kwalitatief onderzoek gingen publiceren is hetzelfde patroon zichtbaar. Het gaat vrijwel altijd om interviewonderzoeken. Deze oververtegenwoordiging van een bepaald soort waarneming leidt er niet alleen toe dat andere vormen van kwalitatief onderzoek, zoals etnografie en de gevalsbeschrijving, ten onrechte worden ondergewaardeerd; er wordt daarmee een verkeerd beeld geschapen van wat kwalitatief onderzoek is, alsof het slechts zou gaan om een ander soort ondervraging.

Het kernthema van deze bijdrage is dat er in het kwalitatieve onderzoek geen sprake is van enkele specifieke ('kwalitatieve') vormen van waarneming, zoals een focusgroep of half- of niet-gestructureerde interviews, maar eerder van een scala van waarnemingsmethodes die passen bij het onderzoeksobject.

Kenmerken van kwalitatief onderzoek

Wester en Hak hebben kwalitatief onderzoek als volgt omschreven:

Kwalitatief onderzoek is een proces van het stapsgewijs opbouwen van een theorie (begrippen en theoretische relaties daartussen) die betrekking heeft op (delen van) de geleefde wereld, waarin bij iedere stap in dat proces opnieuw het vinden van toegang tot de onderzoch-

te wereld en de interpretatie van de daarbij verzamelde gegevens voortkomen uit een reflectie op het tot dan toe opgebouwde inzicht.⁸

Het uitvoeren van kwalitatief onderzoek is goed te vergelijken met een leerproces waarin de onderzoeker zijn voorlopige ideeën steeds beter afstemt op het onderzoeksveld door waarnemingsprocedures uit te proberen en vraagstellingen toe te spitsen. Perioden van waarneming en analyse wisselen elkaar daarom af; zij worden gestuurd door een voortdurende reflectie op de resultaten.⁹

Hieruit volgt dat de *waarneming* in kwalitatief onderzoek open en flexibel is, zowel wat betreft de keuze van de waarnemingsmethode (ondervraging, observatie) als de manier waarop deze wordt toegepast (meer of minder gestructureerd, met toevoegingen en met weglatingen ten opzichte van eerdere waarnemingen) als naar de aantallen en soorten eenheden waarop deze methodes worden toegepast (steekproef). Wat betreft de waarnemingsmethode kan een onderscheid worden gemaakt tussen enkele bij uitstek 'kwalitatieve' methoden (zoals het open interview, participerende observatie, het maken van foto's, films en video's, het door respondenten laten maken van dagboek aantekeningen, het verzamelen van bestaande documenten) en methoden die ook in niet-kwalitatief onderzoek worden gebruikt (zoals de vragenlijst, het observatieschema of experimentele tests). Terwijl kwantificerend onderzoek zich moet beperken tot in numerieke data omgezette gegevens, kan kwalitatief onderzoek zowel kwalitatieve gegevens als getallen verwerken.

Methodologische criteria

Kwalitatief onderzoek staat en valt bij een bijzondere openheid van geest van de kant van de onderzoeker. Deze openheid is nodig om

Dit artikel is het tweede uit een serie van vier over de methoden van kwalitatief onderzoek. H&W publiceert steeds meer kwalitatief onderzoek. Een heldere beschrijving van de methodologie kan ook gewone lezers helpen om de waarde van dergelijk onderzoek beter in te schatten.

In september verscheen een korte inleiding van Philipsen en Vernooy (H&W 2004;47:454-7); in het novembernummer verschijnt een artikel van Wester over de analyse van data in kwalitatief onderzoek en in het decembernummer gaan Van Zwieten en Willems in op de strategische betekenis van het begrip objectiviteit in het denken over kwalitatief onderzoek.

Auteursgegevens

dr. T. Hak, universitair hoofddocent methodologie, Rotterdam School of Management / Faculteit Bedrijfskunde, Erasmus Universiteit Rotterdam, Postbus 1738, 3000 DR Rotterdam.

Correspondentie: thak@fbk.eur.nl

Mogelijke belangenverstrengeling: niets aangegeven.

flexibel te kunnen reageren op wat zich in waarnemingsituaties voordoet en om het conceptueel kader daaraan aan te passen. Maar hoe moet nu worden vastgesteld of de reacties van de onderzoeker, dat wil zeggen diens beslissingen over wat vervolgens moet worden waargenomen en over hoe dat moet gebeuren, de (meest) juiste zijn? Welke criteria moeten hierbij worden gevolgd? Het uitgangspunt hierbij is dat algemene methodologische principes als geldigheid en betrouwbaarheid ook voor kwalitatief onderzoek opgaan, maar aangepast aan de specifieke kenmerken van dit onderzoek. Daarnaast moeten, evenals bij kwantitatief onderzoek, eisen worden gesteld aan de representativiteit van de waarnemingen. Wat betreft de discussie over de hierbij noodzakelijke aanpassingen (of 'vertalingen') van deze algemene principes heeft in Nederland met name Smaling zich verdienstelijk gemaakt.¹⁰

Geldigheid

Veel doelstellingen en onderzoeksvragen van kwalitatief onderzoek vereisen dat in de onderzoeksgegevens de betekenisverlening in de taal van de betrokkenen tot uitdrukking is gebracht (*the actor's point of view, members knowledge*). Een veel voorkomende, maar verkeerde gevolgtrekking uit dit vereiste is dat we de actoren naar hun betekenisverlening moeten *vragen*. Vandaar de populariteit van het interview en de focusgroep. Dat zijn immers manieren waarmee de onderzoeker op een relatief eenvoudige wijze talige uitingen van de betrokkenen kan verzamelen. In de praktijk is deze voorkeur voor ondervraging – en ook de veel voorkomende opvatting dat de focusgroep een methode van ondervraging is – doorgaans misplaatst.

In de eerste plaats is de onderzoeksvraag meestal niet gericht op betekenisverlening op zichzelf, maar op die betekenisverlening als verklaring voor een bepaald handelen of voor het verloop van een proces. De verbinding tussen het handelen en de betekenis van diezelfde handeling voor de betrokkene staat centraal. De relatie tussen enerzijds de verhalen (in interviews of focusgroepen) van respondenten of informanten over de betekenis die een situatie of een handeling voor hen heeft en anderzijds de werkelijke betekenis daarvan in de handelingssituatie is problematisch. Er zijn vele redenen, waaronder cognitieve en situationele, waarom de respondent niet kan (in de meeste gevallen) of niet wil (soms) zeggen wat die betekenis was of is. Vaak kan de gezochte betekenis helemaal niet worden verwoord, maar moet deze worden afgeleid uit de handelingen zelf of uit de opmerkingen die daarbij al doende worden gemaakt. Daarom is observatie – en het zonnodig onmiddellijk ter plekke ondervragen – doorgaans een veel *geldiger* (of meer valide) waarnemingsmethode dan de ondervraging in een individueel interview of groepssituatie.

Nem als voorbeeld een onderzoek van Freeman en Sweeney naar de redenen en omstandigheden waarin huisartsen geen gebruik maken van de evidence die zij wel kennen (zie *kader op p. 504*).¹¹ Door huisartsen over deze kwestie met elkaar in groepen te laten praten – zoals in dit onderzoek werd gedaan – kan men ongetwijfeld allerlei verhalen hierover horen die men vervolgens kan ana-

De kern

- ▶ Kwalitatief onderzoek bestaat net als ieder ander onderzoek uit het verzamelen en analyseren van gegevens bij een vraagstelling of hypothese.
- ▶ Voor de waarneming in kwalitatief onderzoek gelden in principe dezelfde methodologische criteria als voor ander onderzoek.
- ▶ Kwalitatief onderzoek wordt gekenmerkt door de voortdurende afwisseling van waarneming en analyse.
- ▶ In het soort kwalitatief onderzoek dat in medische tijdschriften het vaakst wordt gepubliceerd worden conclusies getrokken uit een serie eenmalige niet- of half-gestructureerde interviews met een beperkt aantal personen (meestal patiënten) die op min of meer toevallige gronden zijn geselecteerd. Dit is de in methodologisch opzicht meest problematische en doorgaans ook de slechtst uitgevoerde vorm van kwalitatief onderzoek.
- ▶ Waarneming via focusgroepen wordt meestal ingezet voor de verkeerde vraagstellingen en wordt vaak op een onjuiste manier uitgevoerd.
- ▶ Het beste kwalitatieve onderzoek bestaat uit een opeenvolging van verschillende deelprojecten waarbij de doelstelling en vraagstelling (en daarmee de meest adequate waarnemingsmethode) van ieder deelproject steeds afhankelijk zijn van de resultaten van voorgaande deelprojecten.

lyseren. Het voordeel van de groepsdiscussie is bovendien dat daarin – door argumentatie en tegenargumentatie – een schifting kan worden aangebracht tussen betere en minder goede verhalen, althans in de ogen van collega's. Het resultaat van deze procedure is echter materiaal over *hoe huisartsen consensus creëren over wat voor hen als collega-huisartsen aanvaardbare verhalen over het niet gebruiken van evidence zijn*. We hebben op deze manier nog helemaal geen materiaal verzameld over redenen en omstandigheden waarin dit – het niet gebruiken van evidence – plaatsvindt, althans niet direct. Een onderzoeker kan op een meer geldige manier waarnemingen doen die voor de onderzoeksvraagstelling noodzakelijk zijn door bijvoorbeeld (in dit geval) gevallen waar te nemen door middel van participerende observatie en daarbij ter plekke de huisarts te ondervragen, of door gevallen waar te nemen door geluids- of beeldopname en achteraf gedetailleerd daarover na te praten (een *focused interview*, al of niet met *stimulated recall*).^{12,13} Maar die meer geldige waarnemingsmethoden zijn in veel opzichten ingewikkelder. Zij kosten meer tijd en meer praktische rompslomp. De populariteit van de individuele en groepsondervraging in het kwalitatieve onderzoek heeft dan ook veel te maken met het gemak waarmee deze methodes kunnen worden toegepast. Je praat wat met mensen en zie daar!

Een ander voorbeeld ontleen ik aan het etnografisch onderzoek van Sudnow naar sterven in het ziekenhuis.¹⁴ Als Sudnow ziet dat een verpleegkundige de ogen van een stervende dichtdrukt nog voordat deze gestorven is, vraagt hij haar naar de redenen daarvan. Als antwoord op deze vraag legt zij hem uit dat het niet netjes is om dit sterven zijn natuurlijke beloop te laten, omdat dit

Het onderzoek van Freeman en Sweeney

In deze serie over de methodologie van kwalitatief onderzoek wordt in alle artikelen een artikel uit de BMJ als een van de voorbeelden gebruikt. Hier geven we een korte uitleg.

Freeman en Sweeney deden een kwalitatief onderzoek om een antwoord te krijgen op de vraag waarom huisartsen zich niet aan evidence-based richtlijnen houden. Ze hielden 3 focusgroepen van in totaal 19 huisartsen (13 mannen, 6 vrouwen) in het zuidwesten van Engeland. De 3 groepen bestonden uit een mix van stads- en plattelandshuisartsen, afkomstig uit verschillende, geografisch van elkaar gescheiden gebieden. Tijdens de groepsbijeenkomsten presenteerde een van de huisartsen een casus waarin hij de richtlijn niet gevolgd had, hoewel hij deze welkende. De groep discussieerde vervolgens over de redenen waarom de richtlijn niet gevolgd was. Daarbij werd veel aandacht besteed aan de arts-patiëntrelatie en de gevoelens die het consult opriep bij de huisarts. Alle groepsbijeenkomsten werden opgenomen en voor de analyse uitgetypt. De auteurs deden 3 analyses gezamenlijk, de rest individueel. Ze bespraken samen de resultaten van de analyses om gemeenschappelijke thema's vast te stellen.

Uit het onderzoek bleek dat huisartsen positief stonden tegenover evidence-based richtlijnen en die ook vaak implementeerden. Barrières die implementatie verhinderden, waren onder andere de persoonlijke ervaringen van de huisarts, de arts-patiëntrelatie, het verschil tussen eerste en tweede lijn en logistieke problemen. Het implementeren van evidence is niet het resultaat van een eenvoudig lineair proces, maar van een gezamenlijke beslissing van huisarts en patiënt. En daarbij is soms de conclusie dat de regels liever niet toegepast moeten worden.

Freeman AC, Sweeney K. *Why general practitioners do not implement evidence: qualitative study*. BMJ 2001;323:1100-2. (zie www.bmj.com voor het volledige artikel)

zou betekenen dat de collega in de ochtendploeg het lichaam zou moeten afleggen. Het is moeilijk voor te stellen dat dit onderwerp zelf of de specifieke handeling – nog-levenden als doden behandelen – in een interview (los van de context van observatie) aan de orde zou hebben kunnen komen. Nog minder waarschijnlijk is het dat daarbij dan de lokaal – tijdens de handeling – geldende intentie zou worden genoemd, om nog maar te zwijgen over de onbespreekbaarheid hiervan in een groepsdiscussie.¹⁵

Uit deze voorbeelden en het daarmee geïllustreerde methodologische principe volgt het volgende voorschrift:

► Als de vraagstelling betrekking heeft op handelingen – en dat is vaak het geval – dan is observatie (via participatie of via beeld- of geluids opnames) de (meest) geldige waarnemingsmethode.

Het is fout om in deze gevallen de voorkeur te geven aan individuele of focusgroepondervraging. Alleen als het onderzochte gedrag om praktische, juridische of ethische redenen niet kan

worden geobserveerd – zoals in onderzoek naar waarom en wanneer men wel of niet een condoom gebruikt – kan van deze regel worden afgeweken. Maar dat moet dan expliciet een tweede – zij het onvermijdelijke – keuze zijn. Tevens moet in zulke gevallen het ontbreken van directe waarnemingsgegevens zo veel mogelijk worden gecompenseerd door te trachten in de ondervraging zo veel mogelijk gedetailleerde beschrijvingen van gedrag te verkrijgen, voordat er over de redenen, motieven en omstandigheden van die gedragingen kan worden gesproken.

Uit de gevolgde redeneerwijze kan ook een volgend voorschrift worden afgeleid:

► Gebruik individuele ondervraging alleen als waarnemingsmethode wanneer de vraagstelling betrekking heeft op meningen en ervaringen – dat wil zeggen op onderzoeksobjecten waartoe alleen de ondervraagde toegang heeft.

Dit lijkt een vanzelfsprekend voorschrift, maar toch wordt het veelvuldig overtreden. De meeste in medische tijdschriften gepubliceerde kwalitatieve interviewonderzoeken gaan in de eerste plaats over gedragingen – zoals het al of niet volgen van leefstijladvies, of het al of niet innemen van geneesmiddelen – en pas in de tweede plaats over redenen daarvoor. In de meeste van zulke gevallen zou een meer directe waarneming van gedrag – bijvoorbeeld via een dagboekmethode – meer geldige gegevens hebben opgeleverd.

Twee soortgelijke maar inhoudelijk andere voorschriften – gebaseerd op hetzelfde principe, en al enigszins aangestipt in de discussie hierboven – betreffen waarneming via focusgroepen:

► Gebruik focusgroepen nooit als waarnemingsmethode voor individuele meningen en ervaringen. Gebruik de focusgroep alleen als de onderzoeksvraag betrekking heeft op cultureel of sociaal geaccepteerde manieren van spreken.

De ratio voor deze twee, met elkaar samenhangende voorschriften blijkt onmiddellijk als men kijkt naar wat er in een focusgroep feitelijk gebeurt. De focusgroep is een publiek forum waarin de vertrouwelijkheid van de individuele interviewsituatie – en ook die van de anonieme vragenlijst – ontbreekt. Dit maakt het onwaarschijnlijk dat deelnemers zonder (zelf)censuur hun meningen met anderen delen. De discussie in de focusgroep is bij uitstek een vorm van *publieke* discussie waarin duidelijk wordt welke visies of soorten argumenten met betrekking tot het bediscussieerde onderwerp cultureel of sociaal acceptabel zijn. Als die publieke vormen en normen van discussie het object van onderzoek vormen – en dat is het vaak het geval in onderzoek naar (politieke) opinies en in marktonderzoek – dan is de focusgroep een bij uitstek geschikte (en geldige) vorm van waarneming. Vaak wordt de focusgroep echter gezien als een goedkoop middel om van meerdere respondenten tegelijk een mening of ervaring te horen – soms met het argument dat de deelnemers elkaar kunnen helpen hun ervaringen te verwoorden. In zulke gevallen krijgt men echter

een resultaat dat wezenlijk anders is dan wanneer een individueel interview zou zijn gehouden.

Inhoudelijke en ecologische geldigheid

In deze beschouwing heb ik twee soorten van geldigheid samen als maatstaf genomen voor de beoordeling van de juistheid van een waarnemingsmethode bij een gegeven object van onderzoek, namelijk *inhoudelijke* en *ecologische geldigheid*. Inhoudelijke geldigheid slaat op de inhoudelijke passendheid van een waarnemingsmethode bij een onderzoeksobject, zoals de passendheid van observatie als waarnemingsmethode bij gedrag als object en de passendheid van ondervraging in het onderzoek naar gevoelens. Ecologische geldigheid is een vorm van geldigheid waarbij de 'natuurlijkheid' van de waarnemingscontext als criterium dient. Het uitgangspunt van ecologische geldigheid is dat

een mening die een respondent in een interview uit, vaak verschilt van een mening in de 'werkelijke' situatie waarin die mening ertoe doet. Het zoeken naar die mening in een 'werkelijke' context is dan een geldiger waarnemingsmethode. Hetzelfde principe geldt voor de waarneming van gedrag. Gedrag in een experiment of laboratoriumsituatie zal verschillen van die in een werkelijke context. In deze voorbeelden is de keuze voor ondervraging of observatie inhoudelijk even geldig, maar er zijn verschillen in ecologische geldigheid.

Betrouwbaarheid

De voorgaande verhandeling over geldigheid laat zien dat in kwalitatief onderzoek, althans in potentie, grote nadruk ligt op de *aansluiting bij het veld van onderzoek*, welke onder meer wordt nagestreefd door grote zorg te besteden aan het bereiken van voldoende ecologische geldigheid. De vrijheid en creativiteit die van de kwalitatieve onderzoeker worden verlangd komen uitermate van pas bij het zoeken van methodes van waarneming die ecologisch geldig zijn. Maar deze creativiteit staat op gespannen voet met een methodologisch criterium dat in andere benaderingen vaak belangrijker wordt gevonden, namelijk *controleerbaarheid* en *repliceerbaarheid*. Stel dat een creatieve onderzoeker toegang heeft kunnen krijgen tot gedragingen die voor anderen ontoegankelijk bleken (zoals gedragingen in slaapkamers, criminele organisaties, of directiekamers) of tot gedachten die voor anderen niet of moeilijk



Foreign correspondent, 1987 (*w/c on paper*), Dean Graham. Williamson Art Gallery and Museum, Birkenhead, UK/The Bridgeman Art Library

toegankelijk zijn (zoals wanneer een interviewer vanwege diens persoonlijke kwaliteiten iemand gedachten kan laten verwoorden die in andere gespreksituaties nooit gezegd zouden zijn). Als een andere onderzoeker al toegang zou kunnen krijgen tot (in principe) dezelfde gedragingen of gedachten, zou de persoonlijke manier waarop die toegang is verkregen niet leiden tot substantieel andere waarnemingen? Met andere woorden, is de repliceerbaarheid en daarmee de wetenschappelijke waarde van de waarnemingen niet onherstelbaar geschaad?

De *betrouwbaarheid* van het kwalitatieve onderzoek wordt over het algemeen – en door de bank genomen terecht – als laag gezien. In veel discussies over methodes van waarneming in kwalitatief onderzoek wordt deze geringe betrouwbaarheid als een gegeven beschouwd, en wordt dit aanvaarde feit vervolgens voorgesteld als de prijs die noodzakelijk betaald moet worden voor de winst die gemaakt is op het terrein van geldigheid. Ik ben het in hoofdlijnen met deze benadering eens, maar ik zie de geringere betrouwbaarheid van de waarnemingsmethodes die vaak door kwalitatieve onderzoekers worden gebruikt niet als iets om je bij neer te leggen, maar juist als een uitdaging om te zoeken naar werkwijzen die de betrouwbaarheid kunnen verhogen. Ik zal hier een aantal manieren noemen waarmee de betrouwbaarheid van de waarnemingen in kwalitatief onderzoek kan worden vergroot.

Vergroting van betrouwbaarheid

Om te beginnen, er zijn wel degelijk ook onderzoeksobjecten waarvan de waarnemingsmethode met de grootste ecologische geldigheid ook de meest betrouwbare is. Het meest voor de hand liggende voorbeeld is dat van de niet-intrusieve waarneming, waarbij de onderzoeker gebruikmaakt van documenten, beelden, geluidsopnames of andere artefacten die een gedrag of situatie vastleggen zonder enige interventie van zijn of haar kant.¹⁶

Ten tweede kan de betrouwbaarheid worden vergroot door bij meerdere waarnemingen steeds zo veel mogelijk op dezelfde (zo min mogelijk intrusieve) manier te werk te gaan. Dit principe kan worden toegepast op vele aspecten van het onderzoek: de manier waarop het onderzoek bij deelnemers (respondenten, geobserveerden, enzovoorts) wordt geïntroduceerd; de manier waarop waarnemingen worden gedaan of vragen worden gesteld; de volgorde waarin waarnemingen worden gedaan en vragen worden gesteld. Wester gaat hier volgende maand in zijn artikel verder op in.¹⁷

Ten derde kan de betrouwbaarheid worden vergroot door de gegevens automatisch te registreren en niet (alleen) via *real time* aantekeningen van de onderzoeker. Met andere woorden, het is altijd aan te bevelen om observaties en ondervragingen op beeld- of tenminste geluidsband vast te leggen.

Bestrijding van onderzoekersbias

Deze manieren om de betrouwbaarheid te vergroten berusten op vormen van mechanisering of standaardisering, waardoor de waarnemingen van een onderzoeker minder afhankelijk worden van toevallige omstandigheden of van eigenschappen van de waarnemer. Maar omdat meestal alle beslissingen in een kwalitatief onderzoek door de ene onderzoeker worden genomen, vormen zij geen voldoende remedie tegen *onderzoekersbias*. Bestrijding van deze *bias* kan op twee verschillende manieren plaatsvinden. Men kan een onderzoeksteam vormen waarin meerdere personen alle soorten waarnemingen en vormen van analyse doen. Door eenvoudige procedures zoals het vergelijken van protocollen of van transcripten van interviews van de verschillende deelnemers kunnen verschillen in aanpak worden opgespoord en kunnen manieren worden ontwikkeld om deze te verkleinen.

De andere vorm waarmee onderzoekersbias kan worden tegengegaan is door anderen niet als mede-onderzoeker in te schakelen, maar als *peer debriefer*. Door geregeld met een collega te spreken over aspecten van het onderzoek kunnen alle aspecten van de kwaliteit van een onderzoek worden bewaakt, waaronder het bevorderen van de geldigheid van waarnemingen en het tegengaan van selectiebias. Maar een *peer debriefer* kan bij uitstek een rol spelen in het tegengaan van onderzoekersbias omdat dit de bedreiging van onderzoekskwaliteit is die de onderzoeker het moeilijkst zelf kan onderkennen.

Representativiteit

Naast het bewaken van de geldigheid van waarnemingen en het

bevorderen van de betrouwbaarheid daarvan is het tegengaan van selectiebias de derde belangrijke methodologische uitdaging bij iedere waarnemingsstrategie. Hierbij gaat het erom dat de aantallen en soorten eenheden waarop de waarnemingsmethodes worden toegepast – zoals het aantal en de soorten observaties, of het aantal en de soorten respondenten – inhoudelijk representatief zijn voor het verschijnsel dat wordt onderzocht. Het gaat niet om een getalsmatig criterium zoals dat geldt bij statistische representativiteit. In kwalitatief onderzoek wil men immers geen uitspraken doen over de mate waarin een verschijnsel voorkomt in een populatie, maar wil men komen tot een beschrijving en daarmee tot begrip van de betekenisverlening van actoren. Representativiteit betekent in dit verband dat al die betekenissen en de relaties daartussen die in de leefwereld van betrokkenen relevant zijn, in het materiaal vertegenwoordigd zijn. Net als voor geldigheid geldt dat deze relevantie en een 'voldoende vertegenwoordiging' niet getalsmatig kunnen worden uitgedrukt. De onderzoeker moet procedures hanteren die de kans vergroten dat het onderzochte verschijnsel van alle voor het onderzoek relevante kanten is bekeken, zonder dat de onderzoeker er ooit absoluut zeker van kan zijn dat dit voldoende is gebeurd.

De makke van een gemakzuchtige steekproef

In de praktijk van het kwalitatieve onderzoek wordt veel gebruikgemaakt van *convenience sampling* ofwel een steekproeftrekking die is gebaseerd op gemak, dat wil zeggen een keuze voor lokaties, situaties, of personen waarbij de onderzoeker zonder veel moeite waarnemingen kan doen. Deze gemakzuchtige praktijk wordt vaak verdedigd met het argument dat er toch geen statistische analyse wordt toegepast en dat het gaat om het *wat* en *hoe*, niet om het *hoe vaak* of *hoe veel*. Dit is echter een foute redenering omdat het, ook wanneer men alleen een verschijnsel wil beschrijven om het te begrijpen, belangrijk is dat men alle relevante vormen en aspecten van dit verschijnsel heeft waargenomen in plaats van alleen een subset waartoe men 'toevallig' toegang heeft gekregen. De term 'toevallig' is hier misleidend omdat het 'toeval' dat gevallen deel laat uitmaken van een gemakssteekproef juist vaak helemaal niet 'toevallig' is in de betekenis van *random*. Dit laatste begrip betekent dat iedere vertegenwoordiger van een verschijnsel (gedrag, gedachte) evenveel kans heeft om in de steekproef terecht te komen, terwijl dat in de gemakssteekproef juist helemaal niet het geval is. Daarin hebben immers alleen gevallen die – meestal om een niet-toevallige reden – al in het gezichts- of sociale veld van de onderzoeker voorkomen een onevenredig grote kans om in de steekproef terecht te komen. Omdat onderzoekers vaak niet weten hoe de gevallen die op hun weg komen zijn voorgeselecteerd door kenmerken van hun eigen netwerk en leefwereld, kunnen zij deze vorm van selectiebias maar moeilijk doorzien.

Random steekproeven ter vermijding van bias

Een radicale manier om selectiebias te voorkomen is *random sampling*. Deze vorm van steekproeftrekking wordt in het algemeen

gezien als tegengesteld aan een kwalitatieve onderzoeksopzet, maar dat is onterecht. In principe geldt voor kwalitatief onderzoek net zo goed als voor ander onderzoek dat selectiebias kan worden bestreden door random steekproeftrekking. Het is echter om drie redenen in de praktijk veel moeilijker uitvoerbaar. Eén reden is dat men in kwalitatief onderzoek vaak met zulke kleine aantallen (respondenten, gevallen) werkt, dat de kans op inclusie van een relatief zeldzame variant van het verschijnsel klein is. Men mist die variant nu niet door selectiebias (hoera) maar door kansvariabiliteit. Een tweede reden is dat men op deze manier grote aantallen van meer voorkomende varianten in de steekproef krijgt, die voor de analyse – die immers kwalitatief en niet getalsmatig is – weinig belang hebben. De derde reden is dat men voor een random steekproeftrekking een steekproefkader nodig heeft. Zulke kaders bestaan wel voor populaties van personen, maar meestal niet voor verschijnselen. Dit neemt niet weg dat het altijd de moeite loont om na te gaan of in het specifieke geval random steekproeftrekking mogelijk is. Dit is met name vaak het geval bij observaties, die meestal naar tijdstip of context kunnen worden gerandomiseerd. Wanneer men bijvoorbeeld een aspect van het reguliere huisartsengesprek wil onderzoeken kan men een verzameling van opnames van spreekurgesprekken aanleggen die naar huisarts en volgnummer (eerste, tweede, enzovoort van de dag) is gerandomiseerd. Of de kwalitatieve onderzoeker zou een random steekproef kunnen trekken uit de bij het NIVEL aanwezige videotapes.

Theoretische steekproeven

Omdat *random* steekproeftrekking om deze praktische redenen vaak niet mogelijk en meestal inefficiënt is, maakt de kwalitatieve onderzoeker die zich niet laat verleiden tot de gemakssteekproef doorgaans gebruik van de 'theoretische' steekproeftrekking. Dit is een methode waarbij de onderzoeker steeds weer probeert vast te stellen wat het onderzoek – dat wil zeggen, zowel de waarnemingen als de analyse die daarmee gelijke tred houdt – tot dusver heeft opgeleverd en welk type waarneming, en met name ook welk soort waarnemingseenheid, de meest productieve bijdrage zou kunnen leveren aan het verdere verloop ervan. Dit is een hachelijke procedure omdat de onderzoeker in principe altijd in het donker tast over wat de volgende waarneming brengen zal, maar is zeker te verkiezen boven een gemakssteekproef. Trouwens, de theoretische selectie van volgende waarnemingseenheden is bij uitstek een procedure die bewaakt en daarmee verbeterd kan worden via *peer debriefing*.

De gevalsstudie en longitudinaal onderzoek

De gevalsbeschrijving wordt in artikelen als deze vaak behandeld naast interviews, focusgroepen en observaties. Tot dusver heb ik het er echter niet over gehad. De reden daarvan is dat de gevalsbeschrijving geen waarnemingsmethode is. Zij kan het best worden gezien als een voorbeeld van een multimethode (of *mixed method*) opzet, waarvan het kenmerk is dat verschillende waarnemingsmethodes naast of achter elkaar worden gebruikt. Zo kan

men in het kader van een gevalsbeschrijving van een ziekenhuisafdeling: observeren, interviews houden met werkers en patiënten, focusgroepen houden, een gestandaardiseerd vragenlijstonderzoek onder deze deelnemers verrichten, documenten bestuderen en cijfermatig materiaal verzamelen dat toch al routinematig op deze afdeling wordt geproduceerd (zoals over opnameduur, of metingen van kwaliteit van leven bij patiënten).

Goed kwalitatief onderzoek bestaat vaak uit een combinatie of – nog beter – een opeenvolging van verschillende deelprojecten, waarbij de doelstelling en vraagstelling (en daarmee de meest adequate waarnemingsmethode) van ieder deelproject steeds afhankelijk zijn van de resultaten van voorgaande deelprojecten. Zo'n multimethodeonderzoek kan een gevalsbeschrijving zijn, bijvoorbeeld van een ziekenhuisafdeling of behandelprogramma, maar dat hoeft niet. Het zou ook heel goed om het onderzoek van een niet aan een locatie of institutie gebonden verschijnsel kunnen gaan, bijvoorbeeld de wachttijd voor een behandeling, of van een subcultuur. De grens tussen een gevalsbeschrijving en een etnografisch onderzoek is in zulke gevallen moeilijk te trekken.

Een ander belangrijk element van de onderzoeksopzet dat in het voorgaande niet aan de orde is geweest betreft de kwestie of een waarneming eenmalig wordt gedaan of wordt herhaald. Het overgrote deel van de kwalitatieve onderzoeken die in medische tijdschriften zijn gepubliceerd, zijn gebaseerd op eenmalige waarneming, dus op een dwarsdoorsnedenopzet. Vaak hebben de onderzoeksonderwerpen een longitudinaal aspect, bijvoorbeeld wanneer men wil onderzoeken hoe patiënten met kanker zich aanpassen aan de verschillende bezoeken die het ziekte- en behandeltraject met zich meebrengt. Vaak worden dan de eenmalige waarnemingen bij patiënten die zich in verschillende fasen van het traject bevinden samengevoegd, alsof de verschillende waarnemingen betrekking hadden op één patiënt. Of patiënten wordt gevraagd over eerdere fasen te rapporteren op basis van hun herinnering. Deze werkwijzen zijn echter niet goed verdedigbaar. In kwantificerend onderzoek is herhaalde meting bij dezelfde personen om een beloop vast te stellen, zeker in medisch onderzoek, gemeengoed. In kwalitatief onderzoek is dat zeker niet het geval, maar het zou het wel moeten zijn.

Voorbeeld van goed kwalitatief onderzoek

De beste kwalitatieve onderzoeken zijn die waarin aan een veelheid van de hier behandelde criteria is voldaan. Tot slot wil ik graag bij wijze van voorbeeld verwijzen naar een onderzoek dat het hier beschreven ideaal dicht benadert, het onderzoek van Gaynor Lloyd-Jones naar de manier waarop eerstejaars medische studenten zich aanpassen aan een curriculum dat is gebaseerd op de principes van probleemgestuurd onderwijs (PGO).^{18,19} Zij maakte slim gebruik van een soort 'natuurlijk experiment' dat te danken was aan het feit dat in haar medische faculteit een nieuw curriculum met PGO-onderwijs werd ingevoerd. De eerstejaars studenten die aan deze nieuwe studie begonnen, konden niet

Abstract

Hak T. Data collection in qualitative research. *Huisarts Wet* 2004;47(11):502-8.

This article discusses the main methods of data collection in qualitative research. Ways of achieving appropriate levels of validity and reliability and of avoiding selection bias are discussed. Because qualitative research is characterised by an iteration between data collection and analysis, it is quite common that multiple types of data collection are used in a single research project, each of which plays a role in a specific phase of the research project.

Unfortunately, the most common research design published in medical journals is the 'cross-sectional qualitative survey', in which conclusions are based on data from one-off unstructured or semi-structured interviews with a limited number of individuals (usually patients), selected in a more or less unsystematic manner ('convenience sample'). This design is rather problematic.

Data collection through focus groups is also a popular method used in medical and health services research. It is often invalid in terms of the research objective and question.

The best qualitative research consists of consecutive parts. Each of these is steered by a research objective and a question that is dependent on the results of the preceding parts. In each part, a method of data collection is chosen that is appropriate for that part's particular research objective and question.

gebruikmaken van advies van een vorige jaargang. De daaropvolgende jaargang kon echter weer wel gebruikmaken van de ervaringen die aan hen werden overgedragen door de eerste groep. Het onderzoek werd opgezet als een gevalsbeschrijving met twee gevallen, namelijk de twee jaargangen studenten. Het enige belangrijke verschil tussen deze twee jaargangen was het al dan niet aanwezig zijn van een eerdere groep met eigen ervaring in dit PGO-curriculum. Andere omstandigheden waren in principe gelijk. Het onderzoek was longitudinaal opgezet, dat wil zeggen dat de studenten in een jaargang het hele studiejaar door werden gevolgd om zo de ontwikkeling van hun studiemethodes te kunnen volgen. In ieder van de twee jaren werden meerdere waarnemingsmethodes gebruikt, steeds met een specifieke doelstelling.

Begonnen werd met participerende observatie om de studenten en hun preoccupaties te leren kennen, en ook later meer gericht om hun studiemethodes te observeren. Deze observaties werden in een later stadium aangevuld met interviews waarin studiemethodes en opinies daarover werden geëxploreerd. De interviews werden later in het studiejaar gevolgd door focusgroepen die waren samengesteld uit nog niet geobserveerde of geïnterviewde studenten, met name om de bevindingen tot dusver te bevestigen. Ten slotte werden aan het eind van het studiejaar de belangrijkste thema's nog eens aan de hele groep voorgelegd via een gestandaardiseerde vragenlijst. Daarbij werden (gelukkig) de eerdere op kwalitatieve waarneming berustende bevindingen bevestigd. Op theoretische gronden en op basis van de bevindingen in de eerste jaargang (die atypisch was voor de socialisatie in een

onderwijssysteem, omdat er in het 'normale' geval een oudere groep is die adviezen over studiemethodes kan geven), werden hypothesen geformuleerd over de tweede jaargang. Sommige daarvan werden bevestigd, andere niet. Dit onderzoek had vooral een theoretische doelstelling, waarin ideeën uit de onderwijssociologie (deels gebaseerd op eerdere onderzoeken over het medisch onderwijs) en de literatuur over PGO werden getoetst. Dit onderzoek demonstreert hoe theorie ontwikkeld en getoetst kan worden in een gevalsbeschrijving.

Het is niet altijd mogelijk om een kwaliteit te bereiken die in dit voorbeeldig onderzoek is bereikt, maar het kan geen kwaad om de lat voor het kwalitatieve onderzoek in de geneeskunde en de gezondheidszorg wat hoger te leggen dan op dit moment gebruikelijk is.

Literatuur

- 1 Britten N. Qualitative Research: Qualitative interviews in medical research. *BMJ* 1995;311:251-3.
- 2 Kitzinger J. Qualitative Research: Introducing focus groups. *BMJ* 1995;311:299-302.
- 3 Mays N, Pope C. Qualitative Research: Observational methods in health care settings. *BMJ* 1995;311:182-4.
- 4 Jones J, Hunter D. Qualitative Research: Consensus methods for medical and health services research. *BMJ* 1995;311:376-80.
- 5 Keen J, Packwood T. Qualitative Research: Case study evaluation. *BMJ* 1995;311:444-6.
- 6 The AM, Hak T, Koëter G, Van der Wal G. Collusion in doctor-patient communication about imminent death: an ethnographic study. *BMJ* 2000;321:1376-81.
- 7 Savage J. Ethnography and health care. *BMJ* 2000;321:1400-2.
- 8 Wester F, Hak T. De methodologie van kwalitatief onderzoek. In: Hak T, Wester F, redactie. *Kwalitatief onderzoek: de praktijk. Waarneming, analyse en reflectie*. Amsterdam: SISWO, 2003:7-17.
- 9 Wester F. *Strategieën voor kwalitatief onderzoek*. Muiderberg: Coutinho, 1995.
- 10 Smaling A. *Methodologische objectiviteit en kwalitatief onderzoek*. Lisse: Swets & Zeitlinger, 1978.
- 11 Freeman AC, Sweeney K. Why general practitioners do not implement evidence: qualitative study. *BMJ* 2001;323:1100-2.
- 12 Merton RK, Fiske M, Kendall PL. *The focused interview. A manual of problems and procedures*. Glencoe (IL): Free Press, 1956.
- 13 Barrows HS. *Stimulated recall (Personalized assessment of clinical reasoning)*. Carbondale (IL): Southern Illinois University Medical School, 2000.
- 14 Sudnow D. *Passing on: the social organization of dying*. Englewood Cliffs (NJ): Prentice-Hall, 1967.
- 15 Hak T. 'Text' and 'con-text': Talk bias in studies of health care work. In: Sarangi S, Roberts C, editors. *Talk, work and the institutional order*. Berlin: Mouton De Gruyter, 1999:427-51.
- 16 Webb EJ, Campbell DT, Schwartz RD, Sechrest L. *Unobtrusive measure. Non-reactive research in the social sciences*. Chicago: Rand McNally, 1966.
- 17 Wester F. Analyse van kwalitatief onderzoeksmateriaal. *Huisarts Wet* 2004 [ter perse].
- 18 Lloyd-Jones G. A multiple case study of the first year student perspective in a medical undergraduate PBL curriculum [Proefschrift]. University of Liverpool, Department of Primary Care, 2002.
- 19 Lloyd-Jones G, Hak T. Self-directed learning and student pragmatism. *Adv Health Sci Educ Theory Pract* 2004;9:61-73.