

Repliek op de commentaren van Schoonenboom en Wester

*Adri Smaling**

Judith Schoonenboom bespreekt naar aanleiding van de eerste alinea van mijn artikel een aantal misverstanden. Maar daarmee wil ze niet zeggen dat ik me aan die misverstanden heb schuldig gemaakt. Daarom beperk ik me hier tot haar vier kanttekeningen bij mijn verhaal, waarmee ik niet wil zeggen dat de misverstanden die zij noemt van geen belang zijn.

De eerste kanttekening gaat in hoofdzaak over het onderscheid tussen statistische generalisatie en variatiedekkende generalisatie (variatiedekking) en in samenhang ermee het onderscheid tussen statistisch steekproeftrekken en systematisch steekproeftrekken. Eerst over de vormen van generalisatie. Ik wil juist wel een duidelijk onderscheid maken tussen statistische generalisatie en variatiedekking. Aangenomen dat een onderzoeker zijn onderzoeksconclusies generaliseerbaar wil maken naar een populatie (of domein) waarvan de onderzochte gevallen (in ruime zin) maar een deel uitmaken, dan gaat het mij erom dat de kwalitatief onderzoeker niet alleen maar vormen van theoretische generalisatie of overdrachtsgeneralisatie ten dienste staan, maar ook een vorm van inductieve generalisatie die toch geen statistische generalisatie is: variatiedekking.

Bij statistische generalisatie wordt een statistisch-representatieve steekproef getrokken. Hierbij spelen variabelen met frequentieverdelingen een rol, en ook het niveau waarop die variabelen gemeten worden, de steekproefgrootte en randomisatie. Het gaat er niet alleen om onbekende, mogelijk storende variabelen te neutraliseren, maar ook om statistisch verantwoorde en significante toetsen uit te kunnen uitvoeren en schattingen te kunnen maken. Er is niks op tegen om bij een dergelijke statistisch-representatieve steekproef kwalitatieve analyses te doen, eventueel gecombineerd met statistische analyses. Alleen maakt de onderzoeker in het kwalitatieve onderzoek geen gebruik van de frequenties enzovoort. Bij variatiedekking zijn al die kwantitatieve aspecten niet aan de orde. Bij variatiedekking gaat het ook om een andere vorm van steekproeftrekken, namelijk doelgericht steekproeftrekken en in het bijzonder iteratief steekproeftrekken. Na het bereiken van een verzadigingspunt wordt dan geen statistische representativiteit van de steekproef bereikt, maar een variatiedekkende representativiteit.

De verschillen tussen deze twee vormen van representativiteit en de daarmee verbonden ideeën en procedures zijn te groot om verdoezeld te worden door ze in één bakje te stoppen. Dit blijkt ook nog als we het onderscheid tussen aselekt steekproeftrekken en doelgericht steekproeftrekken, met name iteratief steekproeftrekken, bezien vanuit het oogpunt van de steekproefgrootte. Bij vormen van aselekt steekproeftrekken (dus met randomisatieprocedures) kan de steek-

* Prof. dr. Adri Smaling is emeritus hoogleraar methodeleer aan de Universiteit voor Humanistiek. E-mail: a.smaling@uvh.nl.

proefgrootte van tevoren worden bepaald aan de hand van de populatiegrootte en de kans op statistisch significante resultaten bij beoogde statistische toetsen en gewenste betrouwbaarheidsintervallen bij schattingen. Bij iteratief steekproeftrekken daarentegen blijkt de steekproefgrootte pas achteraf, wanneer een saturatiepunt is bereikt. Bij iteratief steekproeftrekken, dus ook bij theoretisch steekproeftrekken, kan niet van tevoren worden bepaald hoe groot een steekproef moet zijn, omdat je niet van tevoren weet na hoeveel gevallen (waarnemingen, personen, organisaties, enzovoort) een saturatiepunt zal worden bereikt. Bij de planning van een onderzoek moet hiermee rekening worden gehouden: je kunt niet altijd van tevoren weten hoe groot een steekproef moet zijn. De representativiteit van de twee steekproefvormen is dus van heel andere aard.

In het kort: de kwalitatief onderzoeker is, als deze al wil generaliseren, niet alleen maar aangewezen op theoretische of analytische generalisatie of op overdracht (*case-to-case transfer*), zoals zo vaak wordt gesteld. Er is ook een inductieve vorm van generaliseren mogelijk en deze hoeft geen statistische generalisatie te zijn met de daaraan verbonden aselechte steekproeftrekking, maar kan variatiedekking zijn met iteratief steekproeftrekken als een van de meest aangewezen procedures. Het is daarom belangrijk om statistische representativiteit te onderscheiden van variatiedekkende representativiteit.

Dit onderscheid is nog betekenisvoller als je bedenkt dat ook bij de selectie van een typisch of exemplarisch geval je zou kunnen zeggen dat dit geval representatief is voor bepaalde andere gevallen. Representativiteit hoeft dus niet te worden beperkt tot statistische en variatiedekkende representativiteit, zoals Schoonenboom suggereert. Al met al is de term representativiteit niet zo geschikt om statistische en variatiedekkende representativiteit samen te nemen: ze verschillen te sterk en representativiteit kan meer omvatten.

Voor wat de kwestie van statistisch steekproeftrekken tegenover systematisch steekproeftrekken betreft heb ik de discussie die Schoonenboom voert met sommige kwantificerende onderzoekers met belangstelling gevolgd. Het is inderdaad zo dat de soep lang niet zo heet gegeten wordt als die door sommige puristische statistisch georiënteerde onderzoekers wordt opgediend. Dit blijkt onder meer uit het bestaan van wat men noemt robuustheidsstudies. In deze studies wordt nagegaan in hoeverre statistische principes kunnen worden geschonden en regels overtreden zonder dat de resultaten van of conclusies bij een analyse veranderen. Losjes gezegd: men gaat na, bijvoorbeeld via simulatieonderzoek, dat Schoonenboom ook noemt, in hoeverre een analyseprocedure tegen een stootje kan bezien vanuit de uitkomst.

De tweede kanttekening gaat over de functies van generaliseren. Zoals Schoonenboom zegt, heb ik me in mijn artikel beperkt tot het generaliseren van onderzoeksconclusies bij een afgerond onderzoek naar niet-onderzochte gevallen in een beoogd domein. En ja, inderdaad, er zijn meer toepassingen van de term generaliseren. Een voorbeeld is de generaliseerbaarheidstheorie in de psychometrie als opvolger en uitbreiding van de klassieke testtheorie. In de klassieke testtheorie is betrouwbaarheidscoëfficiënt α van Cronbach (*Cronbach's α*) zeer bekend. In de generaliseerbaarheidstheorie worden meer storingsbronnen die in een test-situatie werkzaam kunnen zijn, verrekend. De oude betrouwbaarheidscoëfficiënt

wordt vervangen door de generaliseerbaarheidscoëfficiënt. Maar de generaliseerbaarheidstheorie blijft naast de klassieke testtheorie en de item-responstheorie een van de betrouwbaarheidstheorieën. Het gaat in de generaliseerbaarheidstheorie over de betrouwbaarheid van een meetinstrument, bijvoorbeeld een psychologische test, en niet over het generaliseren van conclusies bij een afgerond onderzoek naar niet-onderzochte gevallen. Het begrip generaliseren heeft dus zelfs een toepassing op het terrein van de betrouwbaarheid van een meetinstrument.

Wat betreft de combinatie van interne en externe validiteit, waarover Schoonenboom schrijft, kan ook door iteratief steekproeftrekken, zonder theorie, na het bereiken van een saturatiepunt de reikwijdte van de gevormde categorieën of de geformuleerde bewering worden onderzocht door in het beoogde domein naar afwijkende en negatieve gevallen te zoeken. Als je deze gevallen na herhaalde zoekpogingen niet aantreft, kun je zeggen dat de onderzoeksconclusies naar het beoogde domein kunnen worden gegeneraliseerd.

De derde kanttekening betreft een vorm van generaliseren die Schoonenboom in mijn artikel mist. Ook hier gaat het om een woordgebruik waarbij generaliseren niet betrekking heeft op zaken buiten een afgerond onderzoek. Wanneer in een gerandomiseerd experiment met een experimentele conditie (E) en een controleconditie (C) er een statistisch significant verschil wordt gevonden tussen een groep studenten onder E en een groep studenten onder C, heeft dat volgens de gebruikelijke statistische analyse betrekking op het verschil tussen de twee gemiddelde scores van de twee groepen op een afhankelijke variabele. Maar dit sluit helemaal niet uit dat een student P in conditie E dezelfde score heeft als een student M in conditie C. Je kunt zeggen dat in de statistische analyse binnen een groep generaliseerd wordt over de studenten heen door – voor de conclusie significant verschil of niet – naar het gemiddelde te kijken. Weliswaar spelen bij deze analyse ook varianties een rol, maar de individuele studenten P en M blijven buiten het zicht. Of dit terecht is of niet hangt onder meer van de probleemstelling van het onderzoek af: wat wil de onderzoeker eigenlijk weten?

In de vierde kanttekening wordt het thema van het bereik van een generalisatie aan de orde gesteld. Ook hier ga ik met Schoonenboom mee als ze meent dat nader onderzoek nodig is. Het is bijvoorbeeld lang niet altijd van tevoren duidelijk waar een populatie of domein eigenlijk precies uit bestaat. En het bereiken van een saturatiepunt is ook al niet zo eenvoudig. Toch zou het helpen om na het bereiken van zo'n saturatiepunt verder te zoeken naar confirmerende en niet-confirmerende gevallen om een domein beter in beeld te brengen. Ook het replicatieve steekproeftrekken is bedoeld om de grenzen van generaliseerbaarheid van een propositie scherper in het vizier te krijgen. Er is natuurlijk wel tijd, geld en mankracht nodig om ook echt door te zoeken.

Fred Wester stelt de vraag in hoeverre allerlei vormen van generaliseren met steekproeftrekken te maken hebben. Als ik me, zoals in mijn artikel, beperk tot generaliseren in de zin van overhevelen van onderzoeksconclusies van onderzochte gevallen (personen, organisaties, situaties, interviews, enzovoort) naar een domein van niet-onderzochte gevallen, hoeft niet per se een steekproef getrokken te zijn, althans niet opzettelijk door de onderzoeker vooraf aan of tijdens de uit-

voering van het onderzoek. Dit is bijvoorbeeld het geval bij overdraagbaarheid die achteraf wordt bekeken door een gebruiker van het onderzoeksrapport. Men spreekt in zo'n geval ook wel van 'natuurlijke generalisatie': de lezer van het onderzoeksrapport gaat als het ware vanzelf wel na waar en wanneer de onderzoeksconclusies nog meer zouden kunnen opgaan. In een andere invulling van generaliseren, bijvoorbeeld in de bovengenoemde zin dat een gemiddelde score van een groep de individuele score verdoezelt, is er ook al geen steekproef getrokken. Daarentegen is er bij het generaliseren in het kader van de eerder aangestipte betrouwbaarheidstheorie, die als generaliseerbaarheidstheorie bekendstaat, wel weer sprake van een steekproef: de vragen die geselecteerd zijn om in een vragenlijst opgenomen te worden, worden gezien als een steekproef uit een universum van alle vragen die ook geselecteerd hadden kunnen worden om in de vragenlijst te worden opgenomen. Generaliseren hoeft niet per se met steekproeftrekken te maken te hebben, zeker niet als het om een vooraf geplande steekproef gaat. Je kunt immers achteraf generaliseren. Omgekeerd zijn er allerlei vormen van steekproeftrekken die vrijwel niets met generaliseren in de hier bedoelde betekenis te maken hebben. Enkele daarvan heb ik in mijn artikel genoemd.

Verder stelt Wester de vraag: 'Samengevat, in welke zin is generaliseren voor iedere kwalitatief onderzoeker relevant?' Ik denk dat iedere onderzoeker die wetenschappelijk verantwoord wil werken, nagedacht moet hebben, vooraf aan en tijdens een onderzoek, over de vraag of en in hoeverre de onderzoeksconclusies ook zouden kunnen of moeten opgaan voor niet-onderzochte gevallen, en voor welke gevallen dan eigenlijk. Wel is er onderzoek waarbij deze generaliseerbaarheidsvraag niet zo relevant is. Wanneer bijvoorbeeld een specifiek onderwijsprogramma, ontwikkeld en ingevoerd in één bepaalde school, moet worden geëvalueerd in de zin dat moet komen vast te staan of bepaalde onderwijsdoelstellingen zijn gehaald met het oog op continuatie of niet, dan gaat het niet over de vraag wat het programma zou doen in een andere school. Dergelijke generalisatievragen zijn dan niet aan de orde.

Wester presenteert het geval dat in een exploratief onderzoek blijkt dat bij 35 geïnterviewde jongvolwassen Volvrijders tevreden Volvrijders meer links stemmen en ontevreden Volvrijders een voorkeur hebben voor tv-series als Dexter. Dat is een intrigerend resultaat. Maar het roept allerlei vragen op. Je kunt niet zomaar aannemen dat hier iets algemeen gevonden is, dat dit resultaat ook geldt buiten die 35 geïnterviewden. Hoe selectief is de onderzoeker geweest? Hebben die tevreden linkse Volvrijders misschien een prima onderhouden oude Volvo van minstens zeven jaar oud en de anderen een nieuw veel te duur exemplaar? En er is nog niets duidelijk over de aard van het gevonden verband. Toevallig? Causaal? Maar hoe dan? Direct of indirect via andere, tussenkomende factoren? Welke dan? Enzovoort, enzovoort. De aard van het verband maakt generaliseerbaarheid meer of minder aannemelijk. Als oefenonderzoek kan het onderzoek onder die Volvrijders best acceptabel zijn. Dat hangt ook af van de taak of opdracht die de onderzoeker had. Maar los daarvan schreeuwt het onderzoek om inbedding in ander vergelijkbaar onderzoek of een theorie over tevredenheid en autobezit. Maar het onderzoekje vraagt vooral om empirisch vervolgonderzoek, bijvoorbeeld met een opzet waarin voor replicatief steekproeftrekken gekozen

wordt of een of andere vorm van analytische inductie, waarbij het erom gaat uit te vinden naar wat voor domein een uitspraak over een causaal verband gegeneraliseerd kan worden en naar wat voor domein niet. Bijvoorbeeld wel naar de achterbuurten en niet naar de chique wijken in Stockholm en Gotenburg.

Hoe dan ook, geen enkele kwalitatief onderzoeker kan zich onttrekken aan de generaliseerbaarheidsvraag met het idee dat deze niet bij kwalitatief onderzoek past. Bijvoorbeeld omdat kwalitatief onderzoek alleen maar en exclusief diepte-onderzoek zou zijn, *research-in-depth*. Dit is te gemakkelijk. Een kwalitatief onderzoeker kan hiervoor kiezen, maar als het even kan graag wel met een goede reden.