

ONLINE ONDERZOEK EEN BETROUWBAAR ALTERNATIEF?



ONLINE ONDERZOEK EEN BETROUWBAAR ALTERNATIEF?

AFSTUDEERSCRIPTIE

HANS NAUTA

In het kader van de studie

Internationale Economie & Economische Geografie

aan de

Universiteit Utrecht

Faculteit Ruimtelijke Wetenschappen

Begeleider

Ton van Rietbergen

Met dank aan

Multiscope (<http://www.multiscope.nl>)

Publicatiedatum

December 2003

Contact

e hans@xoki.com

i www.xoki.com

INHOUDSOPGAVE

VOORWOORD	1
1 INLEIDING.....	2
2 OPKOMST VAN ONLINE ONDERZOEK.....	3
2.1 INLEIDING	3
2.1.1 <i>Wat is het internet</i>	4
2.2 DEFINITIE & AFBAKENING ONLINE ONDERZOEK	5
2.3 ONTWIKKELING VAN STEEKPROEF ONDERZOEK.....	6
2.3.1 <i>Het begin van steekproef onderzoek.</i>	6
2.3.2 <i>Telefonische enquêtes staan onder druk</i>	8
2.3.3 <i>Opkomst online onderzoek</i>	9
2.4 ONLINE ONDERZOEK	11
2.4.1 <i>Voordelen online onderzoek.</i>	11
2.4.2 <i>Nadelen online onderzoek</i>	13
2.5 WILLEKEURIG VERSUS NIET-WILLEKEURIG GETROKKEN STEEKPROEVEN	15
2.5.1 <i>Willekeurig getrokken steekproeven</i>	15
2.5.2 <i>Niet-willekeurig getrokken steekproeven</i>	18
3 BETROUWBAARHEID VAN ONLINE ONDERZOEK.....	21
3.1 VOORKOMEN VAN STEEKPROEF FOUTEN	21
3.1.1 <i>Dekkingsfout</i>	21
3.1.2 <i>Steekproef fout</i>	25
3.1.3 <i>Non-respons bias</i>	29
3.1.4 <i>Meetfouten</i>	32
3.2 WEEGMETHODEN	36
3.2.1 <i>Gestratificeerd wegen</i>	36
3.2.2 <i>'Propensity weighting'</i>	39
3.3 ONTWIKKELING VAN INTERNET PANELS	44
3.3.1 <i>Internetpanels</i>	44
3.3.2 <i>Bruikbaarheid van online panels</i>	48
3.4 BLIK OP DE TOEKOMST.....	50
4 CASESTUDY: DRIE ONDERZOEKEN VERGELEKEN	52
4.1 INLEIDING	52
4.2 ONDERZOEKSOPZET	53
<i>NIPO's Wereldwijd E-commerce Onderzoek.</i>	53
<i>CBS De Digitale Economie</i>	53
<i>Multiscope E-Commerce Onderzoek</i>	54
4.3 RESULTATEN.....	55
4.3.1 <i>Internetgebruik</i>	56
4.3.2 <i>Online winkelen</i>	57
4.3.3 <i>Meeste online gekochte producten</i>	58
4.3.4 <i>Redenen om niet online te kopen</i>	59
4.4 CONCLUSIE	61
5 CONCLUSIE.....	62
LITERATUURLIJST.....	64

Voorwoord

Deze doctoraalscriptie is tot stand gekomen in het kader van de opleiding Internationale Economie & Economische Geografie aan de Universiteit Utrecht. De scriptie neemt kwantitatief steekproef onderzoek, en de methode van dataverzameling in het bijzonder, als onderzoeksonderwerp. Daarmee ontbreekt het aan een direct vakinhoudelijk verband met de studie, maar daar staat tegenover dat de bevindingen van deze scriptie bruikbaar zijn voor alle vakgebieden waar online dataverzameling gebezigd wordt.

De keus voor dit onderwerp komt voort uit mijn interesses, de rol van computers in de samenleving, en internet in het bijzonder, hebben altijd mijn belangstelling genoten. Combineer dat met affiniteit met statistiek, die zich ondermeer gemanifesteerd heeft in menig student-assistentschap op de universiteit en meer recent in mijn huidige baan als kwantitatief onderzoeker bij het Ecorys-NEI, en de keus voor online onderzoek als onderwerp is snel gemaakt.

Ik wil mijn begeleider Ton van Rietbergen bedanken voor het feit dat ik op deze methodologische scriptie kon afstuderen, en tevens voor zijn begeleiding, hulp en geduld gedurende het hele proces. Daarnaast wil ik John Kivit van Multiscope bedanken voor de nuttige informatie en feedback die hij mij vanuit de praktijk van online dataverzameling mee kon geven.

Verder wil ik iedereen bedanken die mij gesteund heeft bij mijn enigszins trage studie voortgang en mij gemotiveerd heeft om door te gaan. In het bijzonder wil ik mijn ouders ervoor bedanken dat zij mij de mogelijkheid hebben gegeven om te studeren en dat zij mijn studie ook altijd ondersteund hebben.

De grootste dank ben ik verschuldigd aan mijn vrouw Anna. Ze is niet alleen de belangrijkste bron van motivatie geweest, ook heeft ze het grote geduld kunnen opbrengen wat nodig was voordat mijn studie tot een goed eind gebracht kon worden.

1 Inleiding

De revolutie in informatietechnologie heeft aan het eind van de twintigste eeuw geleid tot innovaties, die een duidelijk zichtbaar effect hebben op het dagelijks leven van mensen. De meest ingrijpende innovatie, die als basis dient voor talloze toepassingen, is de ontwikkeling van “het internet”, een tot de verbeelding sprekend wereldwijd netwerk van computers. Binnen een tijdsbestek van nauwelijks tien jaar is meer dan de helft van de bevolking in veel geïndustrialiseerde landen over een aansluiting op het internet gaan beschikken (NIPO, 2002). Als over enige decennia deze periode gekenschetst zal worden, zal zonder twijfel de snelle opkomst en doorbraak van internet de hoofdrol spelen.

Het scala aan nieuwe mogelijkheden die internet te bieden heeft, is ook van grote invloed op de onderzoekswereld. Er is niet alleen een nieuw onderzoeksgebied ontstaan, een geheel nieuwe methode van enquêteren is mogelijk geworden, door middel van het online afnemen van enquêtes. De vanzelfsprekendheid waarmee voorheen gekozen werd voor telefonische enquêtes, is verdwenen. In toenemende mate positioneert online onderzoek zich als directe concurrent van traditioneel ingericht onderzoek.

Juist doordat online onderzoek op veel aspecten wezenlijk verschilt van traditioneel onderzoek, is het echter nog niet duidelijk in hoeverre de resultaten en betrouwbaarheid van online onderzoek vergelijkbaar zijn met die van traditionele methoden. En dus net zoals meer dan vijftig jaar geleden de grote vraag voor “steekproefonderzoekers” was, of telefonische enquêtes vergelijkbaar waren in resultaten en kwaliteit met persoonlijk afgenomen enquêtes, is het nu de vraag in hoeverre online enquêteresultaten vergelijkbaar zijn met enquêtes afgenomen via de telefoon of de post.

De probleemstelling van deze scriptie luidt als volgt.

In hoeverre kunnen door middel van online onderzoek kwalitatief goede, betrouwbare en representatieve gegevens verkregen worden, en in welke mate moet de methodologische aanpak van online onderzoek verschillen van de traditionele onderzoeksmethodiek.

Een uitgebreide literatuurstudie dient het meeste inzicht te verschaffen in deze kwestie.

Daartoe zal eerst de ontwikkeling van online onderzoek besproken worden in hoofdstuk 2. Na een afbakening van het begrip online onderzoek in paragraaf 2.2, komt in paragraaf 2.3 de ontwikkeling van steekproefonderzoek aan bod, van het persoonlijk interview tot de online enquête. Vervolgens wordt in paragraaf 2.4 een overzicht geschetst van veelgenoemde voor- en nadelen van online onderzoek, en wordt in paragraaf 2.5 een onderscheid gemaakt naar verschillende vormen van online onderzoek.

In het volgende hoofdstuk staat de betrouwbaarheid centraal, en wordt aan de hand van mogelijke steekproeffouten duidelijk waarop gelet moeten worden bij online onderzoek. Het corrigeren van enquêteresultaten door middel van weging staat in paragraaf 3.2 centraal, en in paragraaf 3.3 wordt verder ingegaan op de ontwikkeling van internetpanels die inmiddels aardig aan populariteit gewonnen hebben. Hoofdstuk 3 wordt afgesloten met een korte blik op de toekomst.

Aansluitend wordt in hoofdstuk 4 de vergelijking tussen online en traditioneel onderzoek benaderd met behulp van een praktijkvoorbeeld. Een vergelijking is gemaakt tussen drie onderzoeken over hetzelfde onderwerp, waarvan één onderzoek uitgevoerd is door gebruik te maken van een internet panel. Getracht zal worden te laten zien in hoeverre deze andere onderzoeksopzet invloed heeft op de uiteindelijke resultaten.

Tot slot zullen in de conclusie de belangrijkste bevindingen van deze scriptie, met betrekking tot de betrouwbaarheid van online onderzoek, samengevat worden.

2 Opkomst van online onderzoek

2.1 Inleiding

Online onderzoek, ofwel het gebruik van webenquêtes bij het verzamelen van gegevens, is een veelbelovende uitdaging voor steekproefonderzoek in het algemeen. Maar binnen de onderzoekswereld voerden argwaan en scepsis tegenover deze nieuwe methode aanvankelijk de boventoon, die pas later plaats maakten voor enthousiasme en vertrouwen. Terwijl in snel tempo ouderwetse methoden als papieren thuisbezorgde enquêtes en vooral telefoononderzoek vervangen worden door internetonderzoek, moet opgemerkt worden dat in de beginfase de stuwende kracht achter deze ontwikkelingen niet zozeer wetenschappers uit het onderzoeksmethodologisch veld waren, maar computer-programmeurs die al snel hun programmeerqualiteiten aanwenden om software te ontwikkelen voor web-enquêtes. De meer fundamentele vraag hoe betrouwbaar en representatief de resultaten waren lieten zij echter in het midden.

Inmiddels zijn we een paar jaar verder, en hebben alle grote onderzoeksbureaus al enige jaren ervaring opgedaan met online onderzoek. Ook komt er steeds meer methodologisch onderzoek naar online onderzoek. Er verschijnen inmiddels steeds meer publicaties over dit onderwerp. Zeker nu blijkt dat de internetpopulatie steeds meer gaat lijken op de algemene populatie, groeit de kans dat op termijn online onderzoek het telefonisch onderzoek in belangrijke mate zal verdringen (Gouw & Groenland, 2000).

In dit hoofdstuk wordt, aan de hand van literatuur, een overzicht gegeven van online onderzoek. De focus ligt daarbij op de ontwikkeling die online onderzoek heeft doorgemaakt, de voor- en nadelen, en de verschillende vormen van online onderzoek die mogelijk zijn. De betrouwbaarheid van de resultaten van online onderzoek staat in het volgende hoofdstuk centraal.

Alvorens overgegaan wordt tot deze onderwerpen, wordt voor de volledigheid eerst kort de historie van het internet geschetst.

2.1.1 Wat is het internet

De doorbraak van het internet was geen toevallige innovatie, maar een logische consequentie van de ontwikkeling van microprocessors, computers en netwerktechnologieën in de 80'er jaren.

Ten eerste vond de personal computer steeds vaker zijn weg naar consumenten. Tegelijkertijd maakten bedrijven steeds meer gebruik van (interne) netwerken om zo informatiestromen te faciliteren die de organisatie efficiënter moest laten opereren.

Het was dan ook een kwestie van tijd, voordat deze twee ontwikkelingen zich zouden verenigen in het ontstaan van een wereldwijd netwerk van computers. Een bestaande infrastructuur van militaire oorsprong (het Arpanet) bood een goede basis voor de ontwikkeling van een dergelijk netwerk.

Diverse onderdelen ("protocollen") werden gaandeweg ontwikkeld (in chronologische volgorde):

- E-mail: *Elektronische post*. Applicatie die gebruikt wordt om berichten uit te wisselen tussen elektronische postadressen.
- Telnet: programma waarop op afstand andere computers bediend dan wel geraadpleegd kunnen worden. Vooral bekend van toepassing in bibliotheken.
- Newsgroups; *elektronische prikborden* waarop iedereen berichten en reacties daarop kan publiceren. Een van de eerste toepassingen voor Internet.
- IRC, *Internet Relay Chat*. Soort babbelbox waarop gebruikers via het internet synchroon met elkaar kunnen communiceren.
- FTP: afkorting van *File Transfer Protocol*. Maakt overdracht van bestanden mogelijk via internet.
- Gopher; voorloper van het WWW.
- WWW, het *World Wide Web*. Verzameling van miljoenen documenten die op diverse manieren aan elkaar gekoppeld zijn door middel van hyperlinks. Documenten kunnen zowel tekst als grafische elementen bevatten.

Bron: Marsh (1997), Molenaar (1999).

Gopher en Telnet worden inmiddels niet veel meer gebruikt. FTP, IRC en Newsgroups worden nog wel intensief gebruikt, maar het betreft hier een groep gebruikers die te omschrijven is als "computer-hobbyisten".

Voor het grote publiek is de internet vooral bekend om het WWW en E-mail, de twee meest populaire applicaties. Dit kan verklaard worden door het grote nut dat aan deze applicaties toegeschreven wordt. WWW is de meest gebruiksvriendelijke manier om informatie te delen met grote groepen gebruikers, terwijl bij één-op-één communicatie e-mail een belangrijke functie vervult als vervanging van traditionele post, telefoon of face-to-face contact.

De ontwikkeling van het WWW is doorslaggevend geweest bij de acceptatie van internet bij het grote publiek. Pas toen begin jaren negentig Netscape en een paar jaar later Microsoft met een browser kwamen die een zeer gebruikersvriendelijke toegang tot het wereldwijde web mogelijk maakte, kwam de grote doorbraak. Microsoft zag op tijd dat Netscape enorm succesvolle toepassingen mogelijk maakte voor particulieren, en besloot vervolgens om de slag om de "browser" niet te missen en haar browser 'explorer' gratis weg te geven bij het Windows besturingssysteem. Dientengevolge had nu iedere computergebruiker direct de mogelijkheid om toegang te krijgen tot dit web.

2.2 Definitie & afbakening online onderzoek

Wanneer de in deze scriptie gehanteerde term "online onderzoek" nader bekeken wordt, valt op dat er voor het woord "online" geen duidelijke Nederlandse vertaling bestaat. En ook in het Engelse gedeelte van een Engels-Nederlands Prismawoordenboek uit 1991 staat online nog niet vermeld. Zonder in etymologische feiten te duiken, is wel duidelijk dat het woord "online" pas recent populair geworden is, eerst met name onder computergebruikers. Had de betekenis aanvankelijk nog betrekking op het verbonden zijn via de computer met een andere computer (of netwerk), met de intrede van het internet was het woord "online" al snel synoniem voor het hebben danwel gebruiken van een verbinding met het internet.

Hoewel de term "onderzoek" door zijn generieke bekendheid geen toelichting behoeft, dient toch aangetekend te worden dat in deze scriptie kwantitatief onderzoek besproken wordt, waarmee kwalitatief onderzoek uitgesloten wordt. En hoewel een literatuur-studie waarbij internetbronnen gebruikt worden, ook als "online onderzoek" gezien kan worden, ligt deze vorm van onderzoek ook buiten het onderwerp van deze scriptie.

Door de samenvoeging van de termen *online* en *onderzoek*, wordt in deze scriptie verondersteld dat het kwantitatief onderzoek betreft, waarbij via het internet afgenomen enquêtes gebruikt worden om gegevens op te leveren die geanalyseerd kunnen worden. Het gaat hier dus om kwantitatief steekproefonderzoek, waarbij middels een steekproef informatie verzameld wordt van respondenten, die via het internet geënquêteerd worden. Met nadruk wordt hier over enquêtes gesproken, en wordt het gebruik van de term "interview" vermeden. Het gebruik van de term "interview" wordt in het algemeen geassocieerd met kwalitatief onderzoek waarbij een interviewer in een persoonlijk gesprek informatie vergaart zonder daarbij een standaard vragenlijst af te werken. Bij enquêtes die via het internet plaatsvinden, betreft het vrijwel altijd zogenaamde "*self-administered*" enquêtes, waarbij een enquête met standaardvragen zelfstandig door de respondent ingevuld en verstuurd kan worden.

Misschien zou het beter zijn om van "online dataverzameling" te spreken, aangezien juist bij de dataverzameling de toegevoegde waarde van internet in het onderzoeksproces tot zijn recht komt. Het spreekt echter iets minder tot de verbeelding, vandaar de keuze voor de term "online onderzoek" in deze scriptie, waarbij gekomen is tot de volgende definitie hiervan:

"Kwantitatief onderzoek waarbij gebruikt gemaakt wordt van gegevens die verkregen zijn door enquêtes, die op elektronische wijze via het Internet afgenomen worden"

Juist bij het proces van dataverzameling dient internet een rol te spelen, voordat sprake mag zijn van "online onderzoek". Zo zal een onderzoek, waarbij met de respondenten via de e-mail gecommuniceerd wordt over een af te nemen persoonlijk interview, niet vallen onder de noemer "online onderzoek".

Het onderscheid tussen "online" en "niet-online" onderzoek moet tenslotte ook niet verward worden met het onderscheid tussen elektronische en niet-elektronische enquêtemethoden. Zo valt bijvoorbeeld onderzoek waarbij gebruikt gemaakt is van geretoureerde floppydisks (zogenaamde "disk-by-mail") onder elektronisch maar niet onder online onderzoek. Ook bij telefonische enquêtes wordt al geruime tijd gebruikt gemaakt van geavanceerde software die een groot gedeelte van het hele onderzoeksproces ondersteunt, van het trekken van een steekproef tot de verwerking en analyse van de gegevens.

2.3 Ontwikkeling van steekproef onderzoek

In dit hoofdstuk zal besproken worden hoe kwantitatief steekproefonderzoek van de grond is gekomen, en welke ontwikkeling zij sindsdien doorgemaakt heeft. Gezien het uitgangspunt van deze scriptie, ligt de nadruk daarbij vooral op veranderingen van de methode om enquêtes af te nemen. Het zal blijken dat een breed scala aan ontwikkelingen op zowel technologisch, methodologisch maar ook op cultureel vlak hierbij een rol hebben gespeeld. Bij de ontwikkeling van steekproefonderzoek hebben de Verenigde Staten altijd een voortrekkersrol gespeeld, zowel door de statistici die de methodologische basis legden en verder verfijnden, als door het type massa-consumptie-maatschappij die zich na de oorlog ontwikkelde en waarbij een grote behoefte was aan informatie over consumenten. Mede daarom zal in dit hoofdstuk verschillende keren worden gerefereerd aan de situatie in de Verenigde Staten, maar de ontwikkelingen daar verspreidden zich over het algemeen snel naar Europa en Nederland.

2.3.1 Het begin van steekproef onderzoek.

De gedachte dat op basis van relatief kleine steekproeven uitspraken over een populatie kunnen worden gedaan is nog maar van betrekkelijk recente datum, tenminste wanneer we de geschiedenis van de wiskunde en de natuurwetenschappen als maatstaf nemen. In 1895 hield de directeur van het Noorse Centraal Bureau voor de Statistiek, Andres Kiaer, op het congres van het International Statistical Institute voor het eerst een pleidooi voor “*le dénombrement représentatif*” (Sikkel, 2000). Echt serieus werd hij nog niet genomen, tot dan toe werd immers gesteld dat een steekproef slechts cijfers oplevert die voor de feitelijk waargenomen eenheden gelden en niet voor de populatie. In 1906 vond Kiaer een medestander in Bowley, die met formules kwam voor zekerheidsmarges die gebaseerd waren op standaardafwijkingen. De methode veronderstelde dat elk element in de populatie dezelfde kans had om in de steekproef te vallen.

De echte doorbraak van willekeurig getrokken steekproeven (*probability sampling*) en het bijbehorende kwantitatief onderzoek kwam echter pas na de Tweede Wereldoorlog in de Verenigde Staten, waar bleek dat deze inmiddels betrouwbare onderzoeksaanpak zeer veel kennis kon leveren, die indertijd hard nodig was. De kracht zat hem in het idee dat men de kenmerken van huishoudens, individuen of andere grote eenheden van personen in de maatschappij redelijk nauwkeurig kon schatten, door slechts een paar duizend respondenten te enquêteren. Terwijl de methodologie verder werd verfijnd (met bijvoorbeeld gestratificeerd wegen en cluster methodiek), was er tegelijkertijd steeds meer behoefte aan de informatie door de ontwikkeling van de massa-consumptiemaatschappij.

De te gebruiken enquêtemethode was de persoonlijk af te nemen enquête. Er was indertijd nog geen alternatief, en het was ook niet meer dan normaal aangezien vrijwel alle interacties tussen mensen indertijd plaats vonden via persoonlijke face-to-face contacten.

Gedurende de jaren 60 was de telefoon echter aan een sterke opmars begonnen, zodat in 1970 al 90% van alle Amerikaanse huishoudens over een telefoonaansluiting beschikte, een toename van 100% t.o.v. 1930 (Dillman, 1998). Deze dekking van 90% leek voldoende voor het houden van steekproefonderzoeken via de telefoon. Tot dan toe werden steekproeven altijd getrokken uit een index van woonadressen.

Ook had zich gaandeweg de behoefte ontwikkeld aan onderzoek van populaties die kleiner in omvang waren, hetzij door een lager geografisch schaalniveau (stedelijk of regionaal onderzoek in plaats van landelijk), hetzij door de focus op specifieke doelgroepen in de samenleving. De persoonlijke manier van enquêteren maakte hiervan een kostbare

aangelegenheid, de steekproefgrootte is immers niet direct gerelateerd aan de grootte van de populatie¹. Onderzoeken die op een kleinere populatie betrekking hadden, hadden een steekproefgrootte nodig die bijna identiek was aan grote nationale onderzoeken. Zo was het uitvoeren van dergelijke onderzoeken met behulp van persoonlijk af te nemen enquêtes voor bijvoorbeeld regionale en gemeentelijke instituties een erg kostbare zaak. De behoefte aan een kostenbesparend alternatief voor de persoonlijk af te nemen enquête was dus groot.

Toen het telefoonsysteem eind jaren 60 verder gestandaardiseerd en geautomatiseerd werd, was het bovendien ook mogelijk om het hele land te bereiken zonder tussenkomst van een zogenaamde “operator”. In het begin van de jaren 70 werd de telefoon voor het eerst ingezet voor grootschalig steekproefonderzoek. Na een aanvankelijk sceptische ontvangst, besefte eind jaren zeventig alle methodologen dat de telefonische enquête een goede, betrouwbare en vooral kostenbesparende opvolger was van de persoonlijk af te nemen enquête.

Computers

Met de komst van computers konden gegevens door de telefonisten direct ingevoerd worden in computersystemen, hoewel dit eind jaren 70 nog alleen weggelegd was voor enkele grote en kapitaalkrachtige onderzoeksinstituten. De doorbraak van de personal computer rond 1980 gaf een sterke impuls aan de ontwikkeling van de telefonische enquête, doordat nu ook kleinere bedrijven zich telefonisch marktonderzoek konden veroorloven en gegevens konden invoeren en bewerken met behulp van computers. Decentralisatie en verdere commercialisering van telefonisch (markt)onderzoek was het gevolg.

Gedurende de 80'er jaren werd software ontwikkeld die telefonisch enquêteren verder automatiseerde. De ontwikkeling van zogenaamde Computer-Assisted Telephone Interviewing (CATI) systemen resulteerde in kostenbesparingen en een toenemende efficiëntie van het proces. Zo konden medewerkers constant bellen, zonder tijd te "verliezen" aan het proces van het bellen zelf. Het was dus de telefonische methode van enquêteren die als eerste voordeel behaalde uit de mogelijkheden die computers kunnen bieden

Random Digit Dialing (RDD)

Het random trekken van een steekproef met behulp van de zogenaamde Random Digit Dialing (RDD) methode heeft een zeer prominente rol gespeeld in telefonisch enquête-onderzoek vanaf de jaren 70, en nog steeds dient deze methode vaak als referentie bij de beoordeling van andere steekproeftrekking-methodieken. Daarom volgt hier een korte introductie tot RDD.

Al in 1972 werd een in een artikel van Glasser en Metzger (in Newsweek, 1991) geconstateerd dat een substantieel deel van alle telefoonaansluitingen niet genoteerd waren op openbare lijsten, namelijk 19,3% in 1970. Tevens werd opgemerkt dat de eigenschappen van deze huishoudens met een ongenoteerd telefoonnummer verschilden in vergelijking met huishoudens wier nummer wel genoteerd was in het telefoonboek.

De specifieke numerieke eigenschappen van telefoonnummers maakten echter een manier van telefonische steekproeftrekkingen mogelijk waarbij telefoonlijsten niet als uitgangspunt voor het steekproefkader hoefden te dienen. De lengte en codering van telefoonnummers waren immers bekend, en door het gebruik van technologische hulpmiddelen was het mogelijk om cijfer voor cijfer (digit) een willekeurig (random) nummer samen te stellen. Vervolgens werd dit nummer opgebeld (dialing), en als bleek dat het om een niet bestaand nummer ging, werd het proces herhaald totdat wel een bestaand nummer bereikt werd. Op

¹ De relatie tussen de grootte van de populatie en de steekproef wordt in paragraaf 3.1.2 verder besproken

deze manier kon een nauwkeurige en willekeurige steekproef getrokken worden uit alle bestaande nummers, waarbij inbegrepen ongenoteerde nummers. Door middel van de RDD-technologie werd het zodoende mogelijk om makkelijk een goede willekeurige steekproef van de gehele populatie van huishoudens met telefoonaansluitingen te trekken (Alvarez, 2002).

Het belang van ongenoteerde nummers, die niet vermeld staan in openbare lijsten van telefoonnummers, blijkt uit het feit dat demografische kenmerken vaak verschillen bij respondentgroepen met genoteerde en ongenoteerde nummers. Zo volgt uit een RDD-telefonische enquête naar chronische pijn in Australië van Blyth e.a. (2002) bijvoorbeeld dat groepen respondenten met een ongenoteerd nummer jonger zijn, en meer mannen en kleine huishoudens bevatten dan groepen respondenten met een genoteerd nummer. In datzelfde onderzoek blijkt ook dat het responspercentage, dat met 73,4% aan de hoge kant is, ook voor de ongenoteerde nummers behoorlijk hoog is, namelijk 66,3% tegenover 75% voor de genoteerde nummers (Blyth e.a., 2002). Hierbij zij wel aangetekend dat het karakter en de omvang van deze groep telefoonnummers die niet in openbare lijsten genoteerd staan, verschilt van land tot land, al naar gelang de inrichting van de consumentenmarkt voor telefonie. Soms kan bijna de helft van de telefoonnummers van een land niet genoteerd staan, terwijl in bijvoorbeeld Nederland de meeste nummers standaard vermeld staan in het telefoonboek. Als ongenoteerde telefoonnummers de status van een zogenaamd "geheim nummer" hebben, stijgt daarmee waarschijnlijk ook de kans op een hogere non-respons bij deze nummers.

2.3.2 Telefonische enquêtes staan onder druk

De meest gebruikelijke manier van enquêteren voor grote steekproefonderzoeken gedurende de jaren tachtig en negentig is zonder twijfel de telefonische enquête. Het trekken van steekproeven door het willekeurig selecteren van telefoonnummers stond bijna synoniem voor het juist uitvoeren van een steekproef. De meeste huishoudens hebben een telefoonaansluiting, waardoor het willekeurig trekken van telefoonnummers een routine-aangelegenheid is geworden, zeker met de komst van Random Digit Dialing computer-systemen.

Gedurende de jaren 90 komt deze dominante manier van enquêteren echter steeds meer onder druk te staan. De responspercentages bij telefonisch enquêteren laten een dalende lijn zien, terwijl diezelfde percentages voor bijvoorbeeld papieren enquêtes gelijk gebleven zijn (Gouw & Groenland, 2000). Een groeiende groep moeilijk bereikbare respondenten zet de representativiteit van telefonisch onderzoek verder onder druk.

De rol van de telefoon in de maatschappij blijkt aan verandering onderhevig. Was "telefoon!" voorheen een interruptie die prioriteit verdiende boven andere activiteiten, tegenwoordig wordt telefoon meer als een potentiële spelbreker beschouwd, die uitgesteld of anders effectief gecontroleerd dient te worden. Antwoordapparaten, nummerherkenning en dergelijke technologieën vergroten onze mogelijkheden om zelf te selecteren met wie en wanneer we communiceren. Bovendien blijkt meer en meer communicatie die voorheen via de telefoon verliep, zich verplaatst te hebben naar e-mail (Dillman, 1998).

Ook de dekking van vaste telefoonaansluitingen staat onder druk sinds de doorbraak van de mobiele telefoon. De telefoon wordt steeds meer beschouwd als een persoonlijke toepassing in plaats van een product voor het hele huishouden. Zeker onder de jongere populatie is een vaste telefoonaansluiting steeds meer van secundair belang, op de eerste plaats komt de mobiele telefoon. Daardoor wordt het steeds moeilijker om de hele bevolking te bereiken middels een steekproef onder vaste telefoonaansluitingen.

2.3.3 Opkomst online onderzoek

Aangezien het online onderzoek nog zo kort bestaat, is het nog niet goed mogelijk de opkomst van online onderzoek in historisch perspectief te plaatsen. Verschillende facetten over online onderzoek komen in de andere hoofdstukken aan bod, in deze paragraaf ligt de focus op de rol van de maatschappelijke veranderingen bij de opkomst van online onderzoek. Terwijl midden jaren 90 telefonisch onderzoek steeds meer onder druk kwam te staan, opende de doorbraak van internet een scala aan nieuwe enquêtemogelijkheden, en dat niet alleen voor wetenschappers en marktonderzoekbedrijven.

Iedereen die reeds enige tijd actief is op het internet is ze wel eens tegengekomen, zogenaamde "pop-up" windows waarin gevraagd wordt mee te werken aan een online enquête of poll. Zodra de techniek van het World Wide Web de simpele html-scripting oversteeg, werd al snel gebruik gemaakt van deze mogelijkheden om bijvoorbeeld gegevens te verzamelen van website bezoekers middels een simpele enquête. Tegelijkertijd werd ook een begin gemaakt met het afnemen van online enquêtes voor onderzoeken waar wel een degelijke en in meer of mindere mate wetenschappelijk verantwoord onderzoeksontwerp aan ten grondslag lag. Het is dus belangrijk om dit onderscheid te maken tussen enquêtes die aangemaakt zijn om een website aantrekkelijk voor bezoekers te maken of wat simpele vragen van de website-eigenaar over zijn bezoekers beantwoord te zien, en online enquêtes als deel van een kwantitatief onderzoekproces. Deze laatste vorm van enquêteren staat centraal in deze scriptie.

Wel interessant is dat deze behoefte aan feedback niet op zich staat, het is een trend dat organisaties proberen om allerlei feedback mechanismen in te bouwen in hun procesvoering, teneinde kwaliteitsverbetering te kunnen realiseren (Dillman, 1998). Klantenonderzoek is inmiddels onderdeel van de standaard bedrijfsvoering. De beperkende factor bij het achterhalen van deze informatie is vaak het kostenplaatje. Het behoort niet tot het kernproces van bedrijven, maar tot de overhead. Juist door gebruik te maken van internet bij het afnemen van enquêtes is het mogelijk om de kostprijs van onderzoek per respondent fors te verlagen.

Self-administration versus interview

De verschuiving van enquêtes die door een interviewer (telefoniste) afgenomen worden, naar enquêtes die door de respondent zelf ingevuld kunnen worden, kan in verband gebracht worden met een bredere maatschappelijk trend naar meer "self-administration" (Dillman, 1998). Vele activiteiten die in het verleden interactie met een ander persoon benodigde, zijn nu uit te voeren door het individu zelf, met gebruik van technologisch hulpmiddelen. Geld pinnen, online verlengen van bibliotheek boeken, aandelen kopen, inschrijven voor tentamens, selfservice benzine stations zijn maar enkele voorbeelden hiervan. Het functioneren in de maatschappij vraagt vaardigheden in het volgen van instructies en het werken met apparaten die voorheen niet nodig waren.

Computervaardigheden zijn tegenwoordig onontbeerlijk voor het uitvoeren van het meeste geschoolde werk. Een gebrek aan typevaardigheden en de bekwaamheid om snel geschreven berichten af te leveren beperken in toenemende mate de mogelijkheden en kansen tot slagen in het leven. Dillman (2000, pag. 8) concludeert:

"This renewed emphasis on reading and writing, requisite skills for responding to most self-administered surveys, follows a period in U.S. society in which verbal skills and the reliance on the telephone as a substitute for traditional letter writing became dominant."

Deze hernieuwde nadruk op lezen en schrijven moet echter wel gezien worden vanuit het oogpunt van de dagelijkse handelingen van mensen, en in het bijzonder hoe mensen communiceren *met* computersystemen, en *via* computersystemen met andere individuen. Het feit dat geschreven tekst een meer prominente plaats krijgt (lees: meer gebruikt wordt) in de interactievormen tussen mensen onderling en computersystemen, wil echter niet zeggen dat de kwaliteit van geschreven tekst in het algemeen hierdoor ook toeneemt. In tegendeel, het vluchtige karakter van teksten gepubliceerd op het internet lijkt eerder te leiden tot een toenemende nonchalance en slordigheid.

2.4 Online Onderzoek

2.4.1 Voordelen online onderzoek.

Online onderzoek verschilt fundamenteel van traditionele onderzoeksmethoden omdat de informatie direct digitaal ingevoerd wordt door de respondenten. Hierdoor kan een veel groter deel van het hele onderzoeksproces geautomatiseerd worden, wat weer zijn weerslag heeft op de kosten, snelheid en mogelijkheden van online onderzoek. In dit hoofdstuk komt een selectie van diverse veelgenoemde voordelen van online onderzoek aan bod, die samenhangen met het proces van dataverzameling.

Lage kosten

Wellicht de belangrijkste en meest doorslaggevende reden voor het succes van online onderzoek, zijn de sterk verbeterde efficiëntie en navenante lagere kosten per respondent (Boyer e.a., 2002 & Crawford e.a., 2002). Automatisering heeft dit voor een groot deel mogelijk gemaakt, maar ook het feit dat het hier "self-administered" enquêtes betreft betekent een aanzienlijke kostenbesparing op kantoorpersoneel of telefonisten.

Hoewel aanvankelijk met het opzetten van een online onderzoek systeem nog wel wat tijd en kosten gemoeid waren (setupkosten), is dat met het verschijnen van gespecialiseerde maar zeer gebruikersvriendelijke software op dit gebied veel minder het geval. Met de hulp van een webmaster en de juiste programmatuur kan een enquête binnen zeer korte tijd operationeel zijn (Topp en Pawloski, 2002). Toch is het de vraag in hoeverre wetenschappers op de hoogte zijn van het bestaan van deze software, getuige het groot aantal artikelen waarbij wel het ontwerp en opzet van de webenquête ter sprake komt zonder dat er melding gemaakt wordt van het bestaan van deze gespecialiseerde software. Kostenoverwegingen kunnen hier natuurlijk ook een rol spelen.

Voorts merkt Kim (2003) op dat de kosten bij online onderzoek veel minder gerelateerd zijn aan de schaal (lees: het aantal respondenten) van het onderzoek. De totale kosten van traditionele onderzoeksmethoden zijn in sterke mate gerelateerd aan respondent-specifieke kosten zoals kosten van de telefonisch enquête, danwel de kosten van verzending en verwerking van schriftelijke enquêtes. Bij online onderzoek maken deze kosten een veel kleiner deel uit van de totale kosten van het onderzoek. Samen met het bestaan van *setup kosten* die relatief hoger zijn dan bij traditionele enquêtes leidt dit er toe, zoals ook Schonlau e.a. (2001) opmerken, dat pas bij een bepaalde schaalgrootte het kostenvoordeel gunstig uitpakt voor online onderzoek.

Gedurende de responsperiode zelf worden alle antwoorden automatisch op de juiste plaats in een database weggeschreven. Ook het proces van op te sturen herinneringen kan inmiddels vergaand geautomatiseerd worden, wat nog een verder efficiëntieverbetering oplevert.

Door de lage totale kosten van online onderzoek is er sprake van een decentralisering of democratisering van online onderzoek (Topp en Pawloski 2002). Dit kan echter ook resulteren in een wildgroei aan onderzoeken met een zwakke methodologische basis en mogelijke slechte representativiteit. Het is dus nog maar de vraag of kwantitatief onderzoek van hoge kwaliteit ook onderhevig is aan dit proces van decentralisering.

Snelheid

Een veel gehoord voordeel van online onderzoek heeft betrekking op de grote snelheid, of korte omlooptijd, die mogelijk is bij online onderzoek (Schonlau e.a., 2001). Bij de

verzending van e-mail treedt nauwelijks tijdverlies op, en de internetgebruikers "leggen" dagelijks hun postbus. Zo bleek bijvoorbeeld meer dan de helft van alle respons bij een online onderzoek onder onderzoeksbureaus binnen 3 dagen al binnen te zijn (Totten, 2002). Voorts zijn er online onderzoeksbureaus actief die na het ontvangen van een ontwerpenquête van de opdrachtgever, binnen 4 tot 6 dagen de enquêtes af kunnen nemen en de resultaten kunnen leveren aan de opdrachtgever. Samen met een snelle set-up die door gebruiksvriendelijke software mogelijk is, is het dus mogelijk om een zeer korte onderzoekscyclus te bereiken. Bij gedegen wetenschappelijk onderzoek is deze zeer hoge snelheid wellicht minder van belang, maar in het bedrijfsleven speelt de snelheid waarin resultaten van marktonderzoek opgeleverd kunnen worden een veel belangrijkere rol. Diverse grote online onderzoeksbureaus adverteren met een doorlooptijd (tijd tussen ontvangst enquête en leveren enquêteresultaten) van 3 tot 5 dagen (Harris Interactive, 2003). Ook is het bijvoorbeeld niet ongebruikelijk bij professionele workshops dat de cursisten s'avonds online een korte vragenlijst beantwoorden op basis waarvan de volgende dag het programma aangepast kan worden (Topp en Pawloski, 2002).

Gebruiksvriendelijkheid

Op vele aspecten scoren webenquêtes goed op gebruikersvriendelijkheid. Misschien wel het belangrijkste is dat de respondent zelf kan kiezen op welk tijdstip hij antwoordt (Dillman, 1998). Uitgebreide en ingewikkelde *skip* patronen kunnen verwerkt worden in een online enquête zonder dat de respondent daar enig idee van heeft; de respondent krijgt alleen die vragen te zien die van toepassing zijn op zijn of haar situatie. De mogelijkheden die CATI-software (Computer Aided Telephone Interviewing) in de jaren 80 aan telefonistes bood, door vragen die niet van toepassing waren (door eerdere antwoorden) over te slaan, worden nu toegepast op online enquêtes. Voorts kunnen er instructies verwerkt worden voor specifieke vragen in de enquête, die door middel van een *pop-up window* makkelijk aan te roepen zijn door respondenten die dat nodig hebben (Dillman, 2000). Dit resulteert weer in een betere kwaliteit antwoorden.

Nu steeds meer internetgebruikers vertrouwd zijn met de interface van een browser, zorgen radiobuttons en dropdown boxes ook voor een betere controle van antwoorden (Heerwegh & Loosveldt, 2002). En met name bij vragen met meerdere antwoordcategorieën, hebben schriftelijke en online enquêtes een groot voordeel ten opzichte van telefonische enquêtes, aangezien de respondent niet alle antwoorden hoeft te onthouden voordat hij de vraag beantwoordt (Miller en Panjikanan, 2001). Meer aspecten met betrekking tot de gebruikersvriendelijkheid en ontwerp komen overigens ook in paragraaf 3.1.4 aan bod.

De gebruiksvriendelijkheid geldt echter ook voor de onderzoeker als groot voordeel. Het inlezen van de gegevens vanuit de database in verschillende statistische programma's of spreadsheets gaat zeer eenvoudig in zijn werk. Zodra een (software) systeem is ingesteld, kunnen verder ook taken als het verzenden van herinneringen afgehandeld worden door computersoftware. In z'n algemeen lijkt automatisering steeds meer van de "simpele" taken die voorheen met de hand gedaan werden, over te nemen. En dat niet alleen voor een lagere prijs maar vaak ook nauwkeuriger.

Precieze en accurate antwoorden

Doordat de respondent zijn antwoorden direct invoert in het systeem, bestaat er geen kans meer op fouten bij het overtypen of inlezen van antwoorden zoals dat voorheen nog wel eens voorkwam. Ook worden respondenten niet meer beperkt in de ruimte die beschikbaar is voor open vragen, want door middel van scroll-balken hoeft er geen limiet te bestaan. In een

studie naar verschillen in de lengte van open antwoorden tussen web enquêtes en papieren enquêtes (Macroy, 2002) bleken antwoorden in web enquêtes significant langer en van hogere kwaliteit te zijn dan die bij papieren enquêtes.

Door het ontbreken van een interviewer die de enquête afneemt kunnen ook eerlijker antwoorden gegeven worden op vragen waarbij gevoelige informatie (voor de respondent) verstrekt moet worden. Zo is bijvoorbeeld uit onderzoek van Miller en Panjikanan (2001) gebleken dat bij telefonisch onderzoek vaker sociaal wenselijke antwoorden gegeven worden dan bij zogenaamde *self-administrered* onderzoeken waarbij de enquête zonder interviewer afgenomen wordt.

Gebruik meerdere mediatypen mogelijk

Voorals nuttig bij marktonderzoek van producten is de mogelijkheid om afbeeldingen, geluid en zelfs video te integreren in webenquêtes. Hierdoor kunnen tegen lage kosten respondenten op afstand gevraagd worden een mening te geven over bijvoorbeeld het ontwerp van een bepaald nieuw product. Tevens kunnen afbeeldingen gebruikt worden om de precieze bedoeling van een vraag duidelijk te maken, onder het motto "een foto zegt meer dan duizend woorden". Het plaatsen van een juiste foto die niet alleen bijdraagt tot betere maar ook neutraal gevormde antwoorden is hierbij cruciaal.

Voorzichtigheid met betrekking tot het plaatsen van afbeeldingen in online enquêtes dient trouwens altijd betracht te worden door de ontwerper. Enerzijds kunnen afbeeldingen de aandacht afleiden van de werkelijke belangrijke activiteit van de respondent, te weten het invullen van de enquête. Anderzijds leiden grafische elementen vrijwel altijd tot langere laadtijden van online enquêtes. Maar al te dikwijls heeft de ontwerper van een webenquête de beschikking over een snel internet netwerk, en vergeet bij het ontwerpen van de enquête zich te realiseren dat voor veel thuisinternetters, met name modemgebruikers, veel grafische elementen en dus lange laadtijden kunnen leiden tot frustratie of zelfs voortijdige beëindiging van de enquête.

2.4.2 Nadelen online onderzoek

Een overzicht van online onderzoek waarbij alleen voordelen een aparte paragraaf toebedeeld werd, zou een "ge-biased" en onvolledig overzicht zijn. Daarom volgt hier ook een overzicht van enkele veel genoemde nadelen en/of obstakels bij online onderzoek, waarbij opgemerkt zij dat dit geen uitputtend overzicht van nadelen bedoelt te zijn. Aangezien de meeste kwesties uitgebreid worden besproken in andere delen van deze scriptie, zullen ze hier alleen vermeld en pas later uitgebreid behandeld worden.

Generaliseerbaarheid resultaten

De meest besproken zo niet meest controversiële kwestie betreft de generalisatie van resultaten verkregen middels online enquêtes. Dit heeft vooral te maken met het ontbreken van de mogelijkheid om online een willekeurige steekproef te trekken. Veel wetenschappers en statistici staan sceptisch tegenover kwantitatief onderzoek als blijkt dat het niet gebaseerd is op een willekeurig getrokken steekproef. Zij stellen dat, zelfs in geval dat zou blijken dat het "werkt", er toch eerste een degelijke statistische wiskundige theorie nodig is die verklaart hoe en waarom het werkt (Taylor, 1999).

Lage respons

Een ander veelgehoord nadeel van online onderzoek heeft betrekking op de lage respons. De respons bij verschillende vormen van online onderzoek is doorgaans lager dan bij traditionele onderzoeksmethoden (Cook e.a., 2000). Hoewel de respons als zodanig niet maatgevend is voor de betrouwbaarheid van onderzoek, is het doorgaans zo dat een lage respons leidt tot een minder representatieve steekproef. Een treffend voorbeeld van deze lage respons is de respons bij online enquêtes waarbij respondenten geworven worden via webadvertenties (met behulp van zogenaamde *banners*), die zo rond de 1 procent en vaak zelfs lager ligt (Grossnickle & Raskin, 1999). Maar ook de respons op e-mails die verzoeken om deel te nemen aan een enquête laat zich niet vergelijken met respons op enquêtes over de normale post, en staat mede door spam-praktijken in toenemende mate onder druk (Kim, 2003). De responspercentages van internetpanels kunnen daarentegen wel responspercentages van traditionele onderzoeksmethoden evenaren, Manfreda en Vehovar (2002) noteren zelfs percentages van maar liefst 78%, weliswaar bij internetpanels waarbij deelnemers van tevoren via de telefoon geworven zijn.

Technische kennis en apparatuur

Om online enquêtes af te nemen, moet de onderzoeker bekend zijn met en toegang hebben tot apparatuur om online enquêtes mogelijk te maken. Web-servers, database programmatuur en technische expertise zijn nodig om een online enquête op te zetten. Zo moet er bijvoorbeeld rekening gehouden worden met de snelheid, capaciteit en betrouwbaarheid van de web-servers, maar ook gewerkt worden met wachtwoordtechnieken die de integriteit van de respondent waarborgen. Deze en andere technieken benodigen niet alleen tijd en expertise maar ook setup-kosten die aanzienlijk hoger kunnen uitvallen dan bij traditionele enquête-kanalen (Schonlau e.a., 2001).

Toegang van en tot respondenten

Bij de toegang *van* respondenten tot online enquête speelt als eerste een rol dat maar een gedeelte van de bevolking toegang tot internet heeft. Voorts kan de configuratie van de computer en internet verbinding van een respondent zodanig zijn, dat respondenten geen goede toegang tot de enquête hebben en voortijdig moeten afhaken. Een goed doordacht technisch ontwerp kan hierbij veel problemen voorkomen.

Toegang *tot* respondenten heeft met name betrekking op de methode waarop een online steekproef getrokken kan worden. Traditionele methoden gebaseerd op hetzij een telefoonindex hetzij een adressenbestand kunnen op het internet niet voortgezet worden aangezien er geen goede elektronische adressenbestanden bestaan (Kim, 2003). Zoals eerder vermeld komen deze kwesties uitgebreid aan bod in hoofdstuk 3.

Betrouwbaarheid identiteit respondent

Tot slot is het bij online enquêtes moeilijker om de identiteit van de respondent te controleren, in tegenstelling tot telefonische en schriftelijke enquêtes waar een grotere mate van zekerheid bestaat dat de identiteit van de respondent inderdaad gelijk is aan de persoon met het desbetreffende telefoonnummer of adres (Topp & Pawloski, 2002).

2.5 Willekeurig versus niet-willekeurig getrokken steekproeven

Het verschil kennen tussen willekeurig getrokken en niet-willekeurig getrokken steekproeven is van essentieel belang bij het bestuderen van onderzoek. Vooral als de door middel van een steekproef verkregen uitspraken gegeneraliseerd worden naar een grotere populatie, is het van groot belang om te weten hoe de steekproef tot stand is gekomen.

In deze paragraaf zal meer duidelijkheid verschaft worden over het verschil tussen willekeurig en niet-willekeurig getrokken steekproeven. Deze laatste soort (niet-willekeurig steekproeven) hebben vooral onder invloed van internet een grote groei doorgemaakt, maar zoals duidelijk wordt in de volgende paragraaf is het belangrijk de beperkingen van deze manier van steekproef trekken te kennen. Willekeurig getrokken steekproeven zijn beter wetenschappelijk verifieerbaar, maar lenen zich minder makkelijk voor online onderzoek.

2.5.1 Willekeurig getrokken steekproeven

Bij willekeurig getrokken steekproeven identificeert de onderzoeker de populatie en ontwikkelt hij het steekproefkader van waaruit een random steekproef getrokken kan worden. Hierdoor is van tevoren vast te stellen hoe groot de kans is voor elke persoon in een populatie, om getrokken te worden bij de steekproef. In de praktijk betekent dat bijvoorbeeld het toekennen van een nummer aan alle individuen in de populatie, waarna op basis van deze nummers een random steekproef getrokken kan worden. Meestal dient deze kans op trekking voor elk individu even groot te zijn, maar al naar gelang de opzet van het onderzoek kan hier ook van afgeweken worden.

Door de trekkingskansen van alle individuen in de populatie te kennen, is het in ieder geval mogelijk om de steekproeffout² te berekenen. Zo kunnen met een vaststaande betrouwbaarheid generalisaties gemaakt worden over de populatie. Wat betreft willekeurig getrokken steekproeven bij online onderzoek is een indeling gemaakt in steekproeven uit een *gesloten populatie*, steekproeven uit een *algemene populatie* en *pre-recruited panels* (naar Schonlau et al, 2001).

Gesloten populatie

Met een gesloten populatie wordt bedoeld dat er een lijst bestaat van alle individuen in de populatie, een lijst die doorgaans wordt bijgehouden door een organisatie die een vorm van een soort lidmaatschap kent. Zo kan de populatie uit bijvoorbeeld alle geografiestudenten, alle medewerkers van een bedrijf of alle tijdschriftabonnees bestaan. Als een dergelijke lijst bestaat, is het vervolgens vrij eenvoudig om een steekproefkader vast te stellen, van waaruit een willekeurige steekproef getrokken kan worden. Potentiële respondenten kunnen vervolgens via de e-mail aangeschreven worden en doorverwezen worden naar een website waarop de eigenlijke enquête ingevuld kan worden.

Als het echter ontbreekt aan contactgegevens in de vorm van e-mail adressen van de hele populatie, kan er voor gekozen worden om respondenten via een website te werven. In dat geval is een willekeurig getrokken steekproefonderzoek echter alleen mogelijk, als de populatie gedefinieerd wordt als zijnde "bezoekers van de (betreffende) website". Hoewel ongeschikt voor grootschalig marktonderzoek, kan dit type onderzoek wel een goed uitgangspunt zijn bij onderzoeken naar de waardering van online diensten. Zo kan een bedrijf

² De steekproef fout heeft betrekking op de fout die op kan treden omdat maar een deel van de totale populatie geënquêteerd wordt.

dat nader onderzoek verlangt naar de waardering van haar activiteiten en diensten op het internet door haar klanten, heel goed een website-enquête gebruiken. Zolang met de onderzoekspopulatie de website-populatie wordt bedoeld, en zolang dit altijd duidelijk blijft bij het presenteren van resultaten, kan een goed, betrouwbaar en willekeurig getrokken steekproefonderzoek uitgevoerd worden. Bezoek en respons van website-bezoek zijn immers meetbaar, en bezoekers zijn minder anoniem dan soms gedacht wordt³.

Voorbeeld gesloten populatie

De Amerikaanse Air Force Personnel Center enquêteert een deel van haar actieve medewerkers jaarlijks, tot voor kort middels papieren enquêtes via de post. Sinds een paar jaar is echter een systeem opgezet waarmee enquêtes via het internet afgenomen worden. Voor het contact wordt gebruik gemaakt van de e-mail adressen van de werknemers, die bekend zijn want zij worden verstrekt door hun werkgever. Ingevulde enquêtes worden automatisch verwerkt door zelf ontwikkelde software, zodat de gegevens snel beschikbaar zijn voor verdere analyse. De responspercentages liggen rond de 35 tot 40%, ongeveer gelijk aan de papieren enquêtes die voorheen verstuurd werden. Het grootste voordeel van deze omschakeling van papieren naar online enquêtes is een aanzienlijke kostenreductie in de uitvoering van het onderzoek geweest (Schonlau e.a., 2001).

Algemene populaties

Als er geen e-mail lijsten bestaan van de te onderzoeken populatie, die echter wel een algemeen karakter heeft (bijvoorbeeld alle inwoners van Utrecht of alle bezitters van zangvogels als huisdier) wordt het een stuk moeilijker om een willekeurig getrokken steekproef uit te voeren. Het is immers niet mogelijk om aan de hand van contactgegevens van (een groot deel van) de populatie het steekproefkader vast te stellen. Bovendien zijn alternatieve, niet op lijsten gebaseerde, manieren van steekproef trekken⁴ niet beschikbaar voor het internet. In paragraaf 3.1.2 over steekproeffouten wordt verder ingegaan op de mogelijkheden en beperkingen van internet als het gaat om de methodiek bij het trekken van steekproeven..

Op het moment is de enige manier om een willekeurige steekproef te trekken over een algemene populatie, het gebruik maken van conventionele methoden. Conventioneel, in de zin dat beproefde "off-line" middelen zoals de telefoon of post gebruikt worden om potentiële respondenten te benaderen. Vervolgens kunnen respondenten doorverwezen worden naar een website, of kan om e-mail adressen gevraagd worden voor het verdere contact met de respondent.

Aangezien het bereik van internet nochtans niet compleet is, betekent deze onderzoeks aanpak dus ook dat een alternatieve (lees: traditionele) manier van enquêteren beschikbaar moet zijn voor respondenten zonder internet. Als deze mogelijkheid niet kan worden geboden, moet de onderzoeker genoeg nemen met discrepantie tussen het steekproefkader en de populatie (Schonlau et al, 2001). Er bestaat tot slot ook nog de mogelijkheid om de respondent te voorzien van de nodige hardware en software om de enquêtes toch via het internet te kunnen verzenden. Gezien de kosten die hieraan verbonden

³ Zo is het bijvoorbeeld mogelijk om van alle bezoekers van websites met behulp van het zogenaamde ip-adres gegevens over de provider te achterhalen, waarmee vrijwel altijd het land van herkomst vastgesteld kan worden en vaak zelfs de regio/stad.

⁴ Buiten het internet bestaat de methode van de eerder besproken Random Digit Dialing, die immers ook zonder lijsten functioneert.

zijn, loont dit alleen als de respondenten vaker gebruikt kunnen worden voor andere onderzoeken, en dus willen deelnemen aan een zogenaamd *pre-recruited panel*.

Pre-recruited panels

Een *pre-recruited panel* is een groep van potentiële enquêterespondenten, die geworven is door gebruik te maken van een willekeurige steekproefmethode, en waarbij respondenten bereid zijn aan meerdere enquêtes medewerking te verlenen.

Door het trekken van een willekeurige steekproef, meestal door gebruik te maken van de RDD-techniek, worden respondenten uitgenodigd om mee te werken aan meerdere onderzoeken. Daar staat dan wel meestal een bepaalde vergoeding tegenover. Door alleen bij de steekproeftrekking gebruik te maken van "traditionele" middelen zoals telefoon of post, en het enquêteren zelf via het internet te laten plaatsvinden, kan een willekeurig getrokken steekproef toch gebruik maken van de voordelen (met name kostenbesparing) die internet te bieden heeft bij het afnemen van enquêtes. Zoals net besproken moet er echter wel een oplossing gevonden worden voor respondenten in de steekproef die geen beschikking hebben over internet.

In de Verenigde Staten onderscheidt het grote online onderzoeksbureau Knowledge Networks zich door het installeren van speciale hardware (WebTV) bij respondenten die geen beschikking hebben over internet. In hoofdstuk 3.3 wordt verder ingegaan op dit voorbeeld, en op de ontwikkeling van internetpanels in het kader van online onderzoek.

Mixed mode onderzoeken

Het is ook mogelijk dat online enquêtes geïntegreerd worden in *offline* onderzoek. In dat geval hebben respondenten de keus (!) om de enquête online in te vullen in plaats van via de traditionele manier, meestal telefonisch ofwel via de post, te reageren op enquêtes (Alvarez & VanBeselaere, 2002). Op deze manier kan gebruik gemaakt worden van de kostenbesparende en gebruiksvriendelijke voordelen van het online enquêteren, terwijl de traditionele onderzoeksopzet een goede en willekeurig getrokken steekproef kan waarborgen. Tevens wordt door het bestaan van verschillende enquêtemogelijkheden de populatie niet beperkt tot de populatie van internetgebruikers.

Ook Dillman (1998) merkt op dat het combineren van verschillende methodes steeds meer voorkomt, en mede mogelijk gemaakt is door het feit dat computertechnologieën het makkelijker hebben gemaakt om de gegevens van verschillende enquêtebronnen te combineren. Het nadeel is wel dat bij verschillende enquêtemethoden ook verschillende soorten meetfouten kunnen optreden, wat nadelige effecten tot gevolg kan hebben voor de consistentie van de totale gegevensverzameling. Er is tot nu toe nog te weinig bekend over methode-specifieke effecten die die een rol spelen bij het combineren van gegevens verzameld via het internet en via de post of telefoon (Schonlau e.a., 2001).

2.5.2 Niet-willekeurig getrokken steekproeven

Bij niet-willekeurig getrokken steekproeven is het steekproefkader niet bekend, en heeft dus niet iedere persoon in de populatie een bekende kans (en groter dan nul) om opgenomen te worden in een steekproef. Bij deze niet-systematische benadering om respondenten te werven, ook wel *convenience sampling* genoemd, kan dikwijls de potentiële respondent zichzelf in de steekproef selecteren. Hierdoor is het moeilijker om algemene uitspraken te doen naar aanleiding van de antwoorden van respondenten. Zo is bijvoorbeeld in een onderzoek naar milieuvraagstukken, de kans dat personen die betrokken zijn met en/of geïnteresseerd zijn in het milieu zichzelf in de steekproef "selecteren" veel groter dan voor anderen. Een dergelijk onderzoek zal naar alle waarschijnlijkheid de bezorgdheid van mensen in de populatie over milieuvraagstukken overschatten.

Verschillende vormen van niet-willekeurig getrokken steekproeven zijn te onderscheiden binnen het veld van online onderzoek. Naar een typologie van Alvarez et al. (2000) en Schonlau et al. (2001) zullen hier achtereenvolgens onderzoeken die werken met *uncontrolled distribution*, *self-selected website visitors*, en *volunteer panels* onderzoeken behandeld worden.

Uncontrolled distribution

Zoals de term als zegt, bestaat er bij dit type enquêtes op het internet geen controle over de respondenten en kan in principe iedereen op het internet de enquête invullen. Dit type enquête wordt zeer veel aangetroffen op het internet, het betreft hier vaak korte enquêtes of opiniepeilingen met als belangrijkste doel het vermaken van de bezoekers van de website. Het vergroot het idee van interactie met de bezoekers, een voorbeeld hiervan zijn veel nieuws-gerelateerde websites die elk hun dagelijkse "poll" houden. Vrijwel altijd worden de resultaten van dit type enquête alleen op de betreffende site gebruikt, alhoewel de geschreven pers soms ook de resultaten gebruikt voor nieuwsartikelen. Als daarbij de resultaten vervolgens gegeneraliseerd worden naar een populatie die de internetpopulatie overstijgt, is de representativiteit meestal zeer laag. Of in andere woorden, "the message has been always the same: these are simply not reliable methods" (Vehovar 2002, p. 1).

Dit type enquête is *uncontrolled* omdat in principe iedereen op het internet deel kan nemen. Ook is het mogelijk dat internetgebruikers vaker dan één keer deel nemen aan de enquête. Er zijn manieren om dit te voorkomen, maar slimmere computergebruikers kunnen dergelijke beveiligingen vaak ook weer omzeilen. En zelfs als het mogelijk is om meervoudig gebruik te voorkomen, ontbreekt nog steeds enige controle over de respondenten in de steekproef (Schonlau et al., 2001).

Deze methode van enquêteren brengt dus vele beperkingen met zich mee, en daarom wordt ze doorgaans niet gebruikt in wetenschappelijke onderzoeken. In specifieke gevallen kan deze enquêtemethode echter wel een uitkomst bieden voor onderzoekers. Het gaat dan om onderzoeken waarbij de doelpopulatie moeilijk te bereiken is via andere kanalen, terwijl de absolute cijfers of de generaliseerbaarheid van die cijfers van ondergeschikt belang zijn in de onderzoeksopzet ten opzichte van de uitkomsten in het algemeen.

Men kan bijvoorbeeld denken aan een onderzoek naar de slachtoffers van huiselijk geweld. Schonlau et al. (2001) draagt in dit kader een mooi voorbeeld aan van een onderzoek van Coomber naar het versnijden van harddrugs door drugsdealers over de hele wereld. Meer specifiek wilde hij weten hoe gebruikelijk het was om drugs te versnijden op de verschillende internationale markten. Aangezien het hier uiteraard onmogelijk was om een steekproefkader vast te stellen, zette hij advertenties op nieuwsgroepen en verstuurde hij emails naar bezoekers van deze nieuwsgroepen die zich bezighielden met harddrugs. Hij riep

potentiële respondenten op om op publieke computers, zoals in bibliotheken, de webenquête in te vullen of om een uit te printen formulier anoniem op te sturen via de normale post. Hij ontving uiteindelijk 80 enquêtes vanuit 14 verschillende landen.

Self-selecting website visitors

Daarnaast zijn er "self-selecting" onderzoeken waarbij potentiële respondenten via één of meerdere websites medewerking gevraagd wordt aan het onderzoek deel te nemen, maar waar in tegenstelling tot voorgaand type er wel sprake is van een systematische voorselectie van potentiële respondenten. Meer concreet komt dat er meestal op neer dat alleen elke zoveelste bezoeker (bijvoorbeeld elke tiende bezoeker) van een website uitgenodigd wordt door middel van een "pop-up" om deel te nemen aan de enquête. Als de populatie vervolgens gedefinieerd wordt als "de bezoekers van de website", kan gesproken worden over een willekeurig getrokken steekproef. Voor alle anders gedefinieerde populaties betreft het hier echter ook een niet-willekeurig getrokken steekproef, aangezien de overige nadelige effecten van *self-selection* blijven bestaan en algemenering van uitkomsten naar een grotere populatie alleen plaats kan vinden bij een steekproef waarbij alle potentiële respondenten in de populatie een bekende kans hebben om in de steekproef getrokken te worden.

Volunteer panel

De methode waarbij een *volunteer panel* gebruikt wordt, berust op het verzamelen van een groep individuen die zich op vrijwillige basis beschikbaar stelt om in de toekomst deel te nemen aan online enquêtes. Ze worden geworven door gebruik te maken van online marketing technieken zoals bijvoorbeeld *banner* reclames en "*pop-ups*" op diverse website's. Bij registratie moeten ze uitgebreid hun profiel invullen, die de beheerder van het panel nodig heeft voor het selecteren en wegen van panelleden voor diverse onderzoeken.

Er zijn verschillende grote marktonderzoekbedrijven die gebruik maken van volunteer panels. In Nederland beschikt Multiscope over een panel van ruim 100.000 individuen, terwijl Bloomerice in samenwerking met Blauw Research een panel van zo'n 200.000 deelnemers beheert. In de Verenigde Staten heeft Harris Interactive een volunteer panel met een database van enkele miljoenen vrijwillige potentiële deelnemers aan enquêtes. De laatst genoemde twee bedrijven maken gebruik van de statistische techniek *propensity scoring* om te corrigeren voor de selectie bias om zo de resultaten te kunnen generaliseren naar een grotere populatie. Deze weegmethode wordt uitgebreid beschreven in paragraaf 3.2.2. Hier volstaat op te merken dat de bewering dat *propensity scoring* succesvol de selectie bias in *volunteer panels* kan corrigeren controversieel is onder wetenschappers (Schonlau et al., 2001).

Hoewel volunteer panels niet gebaseerd zijn op een willekeurig getrokken steekproef, zijn ze wel meer dan andere typen niet-willekeurig getrokken steekproeven in staat om een representatieve steekproef te benaderen (Alvarez, 2000).

Tot slot

Hoewel er dus veel kanttekeningen geplaatst kunnen worden bij het generaliseren van resultaten van een niet-willekeurig getrokken steekproef onderzoek op het internet naar een grotere populatie, zijn er toch situaties mogelijk waarbij een niet-willekeurig getrokken steekproef het juiste instrument is om tot interessante en wetenschappelijk verantwoorde resultaten te komen.

Zo zijn er experimenten waarbij geen behoefte is aan generalisaties over een populatie, maar waarbij binnen een steekproef met behulp van controlevariabelen die weinig correleren met het onderscheid tussen de verschillende groepen respondenten en non-respons, statistische tests uitgevoerd kunnen worden die een groot aantal respondenten nodig hebben (Alvarez et al., 2000).

Bepaalde populaties, die moeilijk bereikbaar zijn via traditionele kanalen, kunnen door een *convenience sample* soms wel of beter bereikt worden, zoals het eerder besproken voorbeeld over het onderzoek naar drugsdealers laat zien. Ook is het mogelijk om een snelle en goedkope *convenience sample* te gebruiken in de beginfase van een onderzoek, als hulp bij het ontwikkelen van hypothesen, definiëren van antwoordcategorieën en achterhalen van andere informatie over het onderzoeksonderwerp (Schonlau et al, 2001).

Vehovar (2000) merkt op dat zelfs zeer simpele self-selecting enquêtes informatie kunnen verschaffen over het proces van *self-selection* zelf. Deze informatie kan met name nuttig zijn voor de beheerders van grote internetpanels. Daar vindt self-selecting werving plaats om een database van potentiële respondenten samen te stellen, terwijl bij de uitvoering van onderzoeken vervolgens door middel van steekproef-quota en post-onderzoek weegmethoden geprobeerd wordt om een goede representativiteit te behalen. Omdat deze manier de willekeurig getrokken steekproefprocedure compleet vervangt, is het des te belangrijker over nauwkeurige informatie te kunnen beschikken wat betreft het proces van self-selection, zodat beter voor eventueel optredende *bias*-en gecorrigeerd kan worden.

3 Betrouwbaarheid van online onderzoek

3.1 Voorkomen van steekproef fouten

Hoewel het gebruik van online enquêtes een grote vlucht heeft genomen, was het uitgangspunt van die meeste onderzoeken niet het terugdringen van steekproeffouten. Het succes van online onderzoek als basis voor wetenschappelijk onderzoek is echter in grote mate afhankelijk van de betrouwbaarheid van de resultaten.

Gelukkig is er vooral de laatste jaren veel methodologisch onderzoek geweest naar de betrouwbaarheid van online onderzoek, waarbij verschillende facetten een rol spelen. Bij het Research on Internet in Slovenia project (www.ris.org) is bijvoorbeeld een indrukwekkende hoeveelheid materiaal verzameld over online onderzoek als methode om gegevens te verzamelen. In dit hoofdstuk zullen verschillende aspecten van online onderzoek aan bod komen, gezien vanuit de invalshoek van de verschillende mogelijke bronnen die kunnen leiden tot steekproeffouten.

Gedurende het steekproefonderzoek zijn er verschillende processen waarbij steekproeffouten op kunnen treden. Ze zijn te rangschikken onder vier verschillende soorten fouten, namelijk de dekkingsfout (*coverage error*), steekproef fout (*sampling error*), non-respons fout en meetfout (*measurement error*). De dekkingsfout treedt op als niet alle personen in de populatie een kans hebben om in de steekproef opgenomen te worden. De steekproeffout heeft te maken met de statistisch vast te stellen fout die ontstaat omdat maar een deel van de populatie geënquêteerd wordt, en is sterk afhankelijk van de steekproefgrootte. Non-respons fouten kunnen op verschillende momenten optreden, en vormen een probleem als de kenmerken van de non-respons respondenten verschillen van kenmerken van de andere respondenten. De meetfout tenslotte heeft betrekking op fouten die voortvloeien uit de interactie tussen de onderzoeker en de respondenten, waarbij met name een juist ontwerp van de enquête van belang is.

Alleen door te streven naar het beperken van al deze vier mogelijke fouten, kunnen betrouwbare uitspraken gedaan worden over de te onderzoeken populatie.

3.1.1 Dekkingsfout

De dekkingsfout heeft betrekking op de mate waarin de populatie waaruit de steekproef getrokken wordt, verschilt van de werkelijk te onderzoeken populatie. Als er met een willekeurig getrokken steekproef gewerkt wordt, bestaat de coverage error uit het verschil tussen de *population of inference*, dat wil zeggen de populatie waarover de onderzoeker uiteindelijke uitspraken wil doen, en de *frame population*, dat deel van de populatie dat geregistreerd kan worden in een steekproef kader (Schonlau e.a., 2001).

Als we bijvoorbeeld geïnteresseerd zijn in het onderzoeken van een populatie van wie we beschikken over alle e-mail adressen, hoeft er in principe geen coverage error op te treden bij de steekproeftrekking. Willen we daarentegen conclusies trekken over alle huishoudens in Nederland, dan kan er bij online onderzoek een forse coverage error optreden aangezien nog lang niet alle huishoudens in Nederland beschikken over internet toegang, en we ook de contactgegevens van veel internetgebruikers niet kennen.

Toch is dat in rap tempo aan het veranderen. De snelheid waarmee de internettechnologie doorgedrongen is in de huiskamers is ongelofelijk. Binnen 4 jaar na de introductie van het World Wide Web, waren er al 50 miljoen gebruikers wereldwijd (Cook e.a., 2000). Als we kijken naar de huidige marktpenetratie van internet in Amerika, blijkt de introductie vijf keer

zo snel te zijn verlopen als de introductie van de telefoon. Verwacht wordt dat in 2010 meer dan 90 procent van de Amerikaanse bevolking toegang tot internet heeft. Dat percentage is vergelijkbaar met het huidige percentage vaste telefoonaansluitingen (Feld, 2001).

Dichter bij huis had in 1998 nog maar 16% van de Nederlandse bevolking van 16 jaar en ouder internettoegang, terwijl eind 2002 dat explosief is toegenomen tot 9,2 miljoen personen, wat neerkomt op 72,6% van de bevolking in deze leeftijdsgroep. In tabel 3.1 is te zien dat volgens dezelfde cijfers van het onderzoeksbureau Nielsen (2003), uitgevoerd in veertig verschillende landen, Nederland na Zweden (85%) de tweede in Europa is. Daarnaast valt de daling in internettoegang in zowel Hong Kong als Australië op, maar deze wordt niet verder toegelicht in de publicatie van het onderzoeksbureau. Opmerkelijk in dit verband is dat juist de Australische dochter van Nielsen in een persbericht naar aanleiding van de cijfers over 2002, geen vergelijking laat zien met de cijfers van 2001 (Nielsen Media, 2003). Het is dus de vraag of deze daling werkelijk heeft plaatsgevonden, of dat ze veroorzaakt wordt door fouten die gemaakt zijn tijdens het onderzoeksproces of bij de dataverzameling.

Tabel 3.1 Percentage personen 16+ met internettoegang (%), ongeachte de locatie

	4 ^{de} kwartaal 2001	4 ^{de} kwartaal 2002
Zweden	84	85
VS	76	79
Nederland	70	73
Australië	76	72
Hong Kong	73	70
Groot-Brittannië	62	68
Duitsland	59	63
Italië	50	56
Frankrijk	52	54
Spanje	32	54
Brazilië	43	45

Bron: Nielsen/Netratings (2003)

Ondanks de aanhoudende groei van huishoudens met internettoegang, is er nog steeds een groot deel van de huishoudens onbereikbaar via internet en heeft internet nog nergens het niveau van de telefoon overtroffen wat betreft dekking.

In de beginjaren van online onderzoek was deze lage dekking van internet één van de grootste obstakels bij het uitvoeren van online huishoudensonderzoeken. Een lage dekking aan zich zou niet zo'n groot probleem vormen als de "off-line" groepen enigszins evenwichtig (lees: random) verdeeld geweest zouden zijn over de verschillende groepen in de samenleving. Het is echter gebleken dat de ontwikkeling van de internetpenetratie alles behalve evenwichtig verlopen is. Zo bestaat er bijvoorbeeld een positieve correlatie tussen internet gebruik en inkomen (Molenaar, 2001). En in een TNO-rapport "Bouwstenen voor het digitale trapveld" (1999) wordt geconstateerd dat allochtonen, ouderen, laag opgeleiden, mensen met een laag inkomen, werklozen en vrouwen (vooral alleenstaande en eenoudergezinnen) achterblijvers zijn op de elektronische snelweg.

Een "digital divide" ?

Tot een jaar of vijf geleden, was een veronderstelde groeiende kloof tussen mensen aangesloten op het informatienetwerk (lees: internet) en diegenen die niet daarop aangesloten waren, het onderwerp van menig debat. Enige bezorgdheid leek ook op zijn plaats, daar

aanvankelijk slechts bepaalde groepen van de samenleving, met name hoog opgeleiden en hoge inkomens, leken te gaan profiteren van een aansluiting op het internet.

Volgens het TNO rapport "Bouwstenen voor het digitale trapveld" (1999) laat de geschiedenis zien dat *early adopters* van innovaties altijd jong, wit, man en hoog opgeleid zijn. Bij de kwalificatie "wit" kunnen vraagtekens gezet worden, want zelfs als daarmee voorheen een onderscheidend kenmerk kon worden aangeduid, lijkt huidskleur in de tegenwoordige tijd geen belangrijke verklarende variabele bij de adoptie van innovaties. Hooguit indirect, via de hoogte van het inkomen, zal een invloed merkbaar zijn.

Bij de overige verklarende variabelen, leeftijd, geslacht, opleiding en wellicht ook inkomen, lijkt het meer waarschijnlijk dat ze samenhangen met de verspreiding van innovaties. De *early adopters* gebruiken als eerste innovaties, en pas na verloop van tijd worden andere groepen van gebruikers bereikt waarvan maar een klein deel als "achterblijvers" kan worden gezien. Dit proces wordt ook wel aangeduid als de S-curve voor de diffusie van technologische innovaties (TNO-STB, 1999).

De ontwikkeling van de internetpenetratie van de laatste jaren laat zien dat we momenteel de zogenaamde "kritische massa" reeds gepasseerd zijn, zowel in de Verenigde Staten als Nederland. Dit betekent dat de verschillen tussen kenmerken van de "online bevolking" en van de totale bevolking niet langer groeien maar juist kleiner aan het worden zijn.

In een analyse van Lynch (2002, p. 5) wordt op basis van een viertal onderzoeksrapportages van de Amerikaanse overheid naar het profiel van de internetgebruikers, tussen 1994 tot 2000, de volgende conclusie getrokken:

"In the six years since the first report the data show rapid increase in the overall level of U.S. digital inclusion and that the inclusion is occurring among most groups regardless of income, education, race or ethnicity, location, age or gender. Groups that were earlier described as digital 'have nots' now were making dramatic gains."

En uit de laatste publicatie van de Amerikaanse overheid (Cooper & Victory, 2002) blijkt dat dit proces van zogenaamde *digital inclusion* zich voortzet. Zo vindt de grootste groei van internet gebruik van 1999 t/m 2001 naar verschillende types huishoudens plaats onder alleenstaande moeders met kinderen (29%). Voor diezelfde periode laat het internetgebruik voor de laagste inkomensklasse een jaarlijks groei van 25% zien, terwijl datzelfde percentage voor de hoogste inkomensklasse op 11% uitkomt. Natuurlijk is het uitgangspunt van de hoogste inkomensklasse anders, maar de constatering dat het juist achterstandsgroepen zijn die bezig zijn aan een inhaalslag, blijft overeind staan. Sterker nog, het rapport laat zien dat diegenen die de minst traditionele gebruikers van internet zijn, mensen met een laag inkomen, lage opleiding en/of ouderen, behoren tot de *fastest adopters* (Lynch, 2002).

En een van de meest beruchte eigenschappen van het archetype internetgebruiker, namelijk *zijn* geslacht, berust inmiddels ook niet meer op feiten. De "gender" barrière is in Amerika al in 2001 doorbroken; 52% van de internet populatie was vrouw, waarmee het nationale verhoudingen benadert (Feld, 2001).

Gedetailleerde maar enigszins gedateerde gegevens over de Nederlandse internetpopulatie t/m 2000 (CBS, 2002) wijzen overigens in dezelfde richting als de Amerikaanse cijfers; die groepen van de samenleving die voorheen sterk ondervertegenwoordigd waren, zijn bezig aan een inhaalslag in hoog tempo.

De tot de verbeelding sprekende metaforen als "de information have and have nots" of de "information rich and information poor" blijken dus inmiddels veel aan kracht te hebben ingeboet.

Dekking van de telefoon

Bij de bespreking van de vooralsnog gebrekkige dekking van internet inzake de totale populatie wordt meestal de dekking van de telefoon als uitgangspunt genomen. De telefonische enquête staat immers gelijk aan een betrouwbare manier van enquêteren met een groot bereik zo was de gedachte, en behoorde in ieder geval tot voor kort tot het standaard instrument bij marktonderzoekers. De dekking van de telefoon was tot voor kort inderdaad goed te noemen, met in de meeste Westerse landen een dekkingsgraad van meer dan negentig procent.

De laatste 10 jaar zijn marktonderzoekers echter zeer bezorgd geworden over het behoud van de kwaliteit van telefonische enquêtes, nu de respons op telefonisch onderzoek al enige jaren structureel blijft dalen (Taylor, 1999). Verschillende oorzaken liggen hieraan ten grondslag. Zo bestaat er een toegenomen irritatie onder mensen om ongevraagd via de telefoon te worden benaderd, en tevens een verwarring onder het publiek van onderzoek met telemarketing activiteiten. Ook de opkomende individualisering en verzakelijking van de maatschappij draagt er aan bij (Willemsen & Lingeman, 2002).

Nieuwe technologische mogelijkheden, zoals nummerherkenning en antwoordapparaten, maken het mogelijk dat een beller eerst "gescreend" kan worden nog voordat de telefoon opgenomen hoeft te worden. Vroeger was het zo dat de telefoon controle uitoefende op het menselijk gedrag doordat een individu zich verplicht voelde om de telefoon te beantwoorden. Men kon immers nooit van tevoren weten of het een belangrijk telefoontje zou zijn, en buiten dat was het ook een vorm van beleefdheid om de telefoon te beantwoorden. De rollen zijn nu echter omgedraaid, zodat het individu controle kan uitoefenen op door wie en wanneer hij bereikt wil en kan worden. Deze controle wordt ook steeds meer erkend als legitiem recht, en nieuwe regels moeten het individu hierbij ondersteunen.

Zo is in de Verenigde Staten vanuit overheidswege een zogenaamde "bel-me-niet-lijst" ingesteld door middel van een website, waar abonnees hun telefoonnummer(s) kunnen registreren. De lijst maakt het weren van alle ongewenste telefoontjes van *direct marketeers* mogelijk, en in slechts enkele maanden tijd zijn al vijftig miljoen telefoonnummers geregistreerd (Volkskrant, 2003). Op het moment van schrijven is echter nog niet bekend of deze lijst rechtsgeldig is en definitief ingevoerd wordt.

Tenslotte heeft het gebruik van mobiele telefoons een grote vlucht genomen, en die mobiele telefoons nemen steeds vaker de plaats in van de vaste telefoonaansluiting (Taylor, 1999). En wellicht zal internettelefonie in de toekomst de dekkingsgraad van reguliere (vaste) telefoonaansluitingen nog verder onder druk zetten. Het is dus de vraag in hoeverre in de nabije toekomst, onderzoeksmethoden waarbij gebruik gemaakt wordt van telefonische enquêtes als '*best practice*' gezien kunnen worden. Indien er een steeds grotere non-respons optreedt, waarbij de non-respondenten kenmerkende groepen vormen binnen de populatie, is het de vraag of nog steeds gesproken kan worden van een werkelijk willekeurig getrokken steekproef.

In zijn algemeen is de dekkingsfout met betrekking tot online onderzoek een probleem dat wel steeds kleiner wordt, naar alle waarschijnlijkheid zal binnen 5 tot 10 jaar internet de (vaste) telefoonaansluiting overtroffen hebben inzake landelijke dekking (Feld, 2001). Maar hoewel het probleem dan wel relatief afneemt, blijft het altijd een belangrijk punt van aandacht want een volledige internet penetratie zal nooit plaatsvinden. Er zullen altijd groepen zijn in de samenleving die niet bereikt kunnen worden via het internet, en aanvullend onderzoek naar deze groepen is continu van belang, zodat deze groepen geïdentificeerd kunnen worden en meer informatie over hen beschikbaar komt. Met deze informatie kunnen de resultaten van online onderzoek vervolgens verbeterd of gecorrigeerd worden. Het is echter de vraag in hoeverre het mogelijk is en blijft om deze *offline* groepen te

bereiken met andere methoden. Wat betreft telefonisch onderzoek is immers al eerder gebleken dat zowel landelijke dekkingspercentages als responspercentages in toenemende mate onder druk staan. Het is waarschijnlijk dat er altijd een aantal groepen mensen in de samenleving niet bereikt kan worden, ongeacht het gebruikte medium.

Maar terwijl het probleem van de dekking, en met name het huidige verschil tussen online dekking en telefonische dekking, wel kleiner zal worden naarmate de internetpopulatie dichterbij het ideaal van een afspiegelingen van de gehele populatie komt, blijft de vraag *hoe* een representatieve steekproef getrokken kan worden via het internet waarschijnlijk langer voor problemen zorgen. Daarover meer in het volgende paragraaf.

3.1.2 Steekproef fout

De steekproeffout wordt gedefinieerd als de fout die ontstaat doordat niet de hele populatie, maar slechts een deel van de populatie (namelijk de steekproef) wordt onderzocht. Het begrip steekproeffout (*sampling error*) kan overigens verschillend geïnterpreteerd worden. Schonlau e.a. (2001) onderscheiden een *variantie* component en een *bias* component.

De variantie component heeft betrekking op het feit dat er altijd een mate van heterogeniteit in de populatie bestaat die tot een specifieke steekproef leidt. Met andere woorden, de variantie component kenmerkt de verschillende uitkomsten die het resultaat kunnen zijn van het trekken van verschillende steekproeven. Elke nieuw getrokken steekproef zal leiden tot haar eigen schattingsparameter. Doorgaans wordt met de steekproeffout in enge zin alleen deze variantiecomponent bedoeld. Het is dan een fout die precies te kwantificeren is in statistische kansen. Een tweede component die Schonlau e.a. onderscheiden bestaat uit de *bias*, die een systematisch verschil tussen de steekproef kenmerken en kenmerken van de werkelijke populatie aangeeft. Een 100% willekeurig getrokken steekproef heeft in principe geen *bias*. Omdat dit echter zelden plaats kan vinden, is het belangrijk om hier ook de methode van steekproeftrekking en de gevolgen voor de *bias* component te behandelen .

Variantiecomponent

De variantiecomponent van de *sampling error* kwantificeert de onzekerheid die bestaat in kwantitatief onderzoek als gevolg van het trekken van een steekproef die kleiner is dan de totale populatie. Zij hangt uiteraard nauw samen met de grootte van de te trekken steekproef. In tegenstelling tot de andere mogelijke oorzaken van fouten, is deze steekproeffout vrij nauwkeurig vast te stellen.

Zo is het logisch dat bij een onderzoek onder de Nederlandse bevolking naar politieke voorkeur een steekproef van 100 respondenten minder betrouwbare resultaten geeft dan een steekproef van 2.000 respondenten. Dat het vervolgens niet veel uitmaakt of 2.000 of 20.000 respondenten geënquêteerd worden, lijkt op het eerst gezicht echter een stuk minder logisch. Toch varieert de benodigde steekproefgrootte alleen sterk bij onderzoeken met een kleine populatie. Zodra de populatie in de tienduizenden gaat lopen stabiliseert de benodigde steekproefgrootte zich, meestal ergens tussen de 1000 en 2000, al naar gelang overige vooronderstellingen die na het volgende voorbeeld verder besproken worden.

Om deze relatie toe te lichten wordt in tabel 3.2 de relatie tussen populatiegrootte en *sampling error* weergegeven, waarbij de benodigde steekproefgrootte als afhankelijke variabele geldt. Om de relatie tussen steekproefgrootte en steekproeffout te kwantificeren is het nodig om nog twee andere factoren te kennen: het percentage betrouwbaarheid waarmee men uitspraken wenst te doen, en de variantie van de populatie in de variabele waarover het onderzoek uitspraken wil doen. In de hierop volgende tabel wordt een 95%

betrouwbaarheidsniveau verondersteld en een maximum variantie (50/50) op een dichotome variabele (ja/nee). Tot slot wordt met steekproefgrootte de *complete* steekproefgrootte bedoeld, waarbij dus geen rekening gehouden wordt met eventuele non-respons.

De getallen in tabel 3.2 zijn tot stand gekomen door gebruik te maken van de volgende formule:

$$N_s = \frac{(N_p) * (p) * (1 - p)}{(N_p - 1) * (B / C)^2 + (p) * (1 - p)}$$

Waarbij

N_s = totale benodigde steekproefgrootte

N_p = grootte van de populatie

p = deel van de populatie dat verwacht wordt te kiezen voor één van de twee antwoordcategorieën

B = acceptabele grootte van de steekproef fout

C = Z-waarde van de gekozen betrouwbaarheidsmarge; bijvoorbeeld 1.96 voor 95% betrouwbaarheid

Tabel 3.2 Steekproefgrootte nodig bij populatiegrootte en steekproeffout

Populatie grootte	+/- 10% Steekproeffout	+/- 5% Steekproeffout	+/- 3% Steekproeffout
100	49	80	92
200	65	132	169
400	78	196	291
600	83	234	384
800	86	260	458
1.000	88	278	517
2.000	92	322	696
4.000	94	351	843
6.000	95	361	906
8.000	95	367	942
10.000	95	370	956
20.000	96	377	1.013
40.000	96	381	1.040
100.000	96	383	1.056
1.000.000	96	384	1.066
1.000.000.000	96	384	1.067

Bron: Dillman, 2000

Dus voor een ja/nee vraag met een 50/50 verdeling in een populatie van 800 individuen, voor een betrouwbaarheid van 95% dat de steekproef schatting van deze vraag binnen een 3%-punt interval ligt van de werkelijke waarde voor de populatie, is een steekproefgrootte nodig van 458 respondenten.

Let wel dat deze tabel weergegeven is om de relatie tussen populatie en steekproefgrootte te illustreren, en gezien de specifieke vooronderstellingen die aan de tabel ten grondslag liggen is het geenszins een algemeen toepasbare en geldende tabel. Afgezien van de eerder besproken betrouwbaarheidsmarges en steekproeffout, is met name het *type* parameter dat geschat moet worden van invloed op de steekproefgrootte. Daarbij geldt dat hoe groter de mogelijke variantie in een parameter, hoe groter de steekproefgrootte. In dit voorbeeld hebben we ons beperkt tot een vraag met twee antwoordcategorieën (ja/nee), maar als een berekening gemaakt moet worden van een vraag met meerdere antwoordcategorieën, bestaat er een grotere mogelijke variantie en is dus ook een grotere steekproefgrootte nodig.

Het blijkt dus dat na een bepaalde drempelwaarde de grootte van een populatie weinig uitmaakt voor de grootte van de te trekken steekproef. Andersom wanneer de populatie klein is, moet een relatief groot gedeelte van de populatie geënquêteerd worden om een zelfde mate van precisie in de resultaten te kunnen garanderen, getuige ook het voorbeeld van een populatie van 800 individuen waar een steekproefgrootte van 458 gewenst zou zijn. Als vervolgens ook een belangrijk kenmerk van online onderzoek in ogenschouw genomen wordt, namelijk de lage marginale kosten per respondent, rijst de vraag of het wellicht niet logischer is om bij dergelijk lage aantallen de hele populatie te enquêteren en het trekken van een steekproef achterwege te laten. Dit kan een goed alternatief zijn, mits de grotere groep respondenten die benaderd moet worden er niet toe leidt, dat het responspercentage onder druk komt te staan. De bias die dikwijls het gevolg is van non-respons, kan een zwaar wegende en negatieve invloed uitoefenen op de betrouwbaarheid van een onderzoek waarvoor een grotere steekproefrekking zelden kan compenseren (Dillman, 2000).

Bias component

Een werkelijk willekeurig getrokken steekproef is de beste garantie voor het ontbreken van een *bias* component in de steekproeffout. Niet-willekeurig getrokken steekproeven, ook wel *convenience samples* genoemd, worden over het algemeen geassocieerd met *biased* gegevens omdat de precieze werking van het mechanisme waarmee de steekproef getrokken wordt, niet bekend is.

Voor onderzoeken met een grote populatie, zoals alle huishoudens/personen in een bepaald land, is het uitvoeren van een willekeurig getrokken steekproef op het internet echter erg lastig. Een kernprobleem bij deze onderzoeken is de manier waarop op een random manier een steekproef getrokken kan worden, waarbij alle personen in de populatie een redelijke kans hebben om aangeschreven te worden. Bij traditionele onderzoeksmethoden wordt gebruikt gemaakt van indexen met contactgegevens van de gehele populatie, waarmee met name bedoeld wordt op de registratiebestanden van de telefoonmaatschappij en naam-adres-woonplaats bestanden van de ptt. Bovendien bestaat bij telefonisch onderzoek ook nog de mogelijkheid om zonder deze lijsten toch willekeurig telefoonnummers te kiezen met behulp van RDD- technieken.

Op het internet ontbreekt echter een dergelijke centrale index van alle internetgebruikers van een bepaald land. Er bestaat geen centrale autoriteit of organisatie die internetgebruikers registreert. Integendeel, er bestaat een veelvoud aan verschillende manieren om toegang tot internet te krijgen, waarbinnen weer een veelvoud aan commerciële en niet-commerciële aanbieders opereren. Zelfs afgezien van juridische kwesties met betrekking tot de bescherming van persoonsgegevens, lijkt het in ieder geval in de nabije toekomst onmogelijk om al deze verschillende administraties samen te voegen.

Er bestaan zogenaamde emailgidsen, maar slechts een zeer klein deel van de internetpopulatie is opgenomen in deze bestanden. Er bestaan inmiddels ook talloze algemene mail-adresbestanden, maar meestal betreft het bestanden die via een specifieke website of verkoopkanaal opgebouwd zijn en totaal niet representatief zijn. Tevens is de houdbaarheid van emaildatabases zeer kort, gezien het feit dat mensen vaak van e-mail adres veranderen. Zo blijkt uit een studie naar e-mail gebruik in Amerika dat 32% van de internet gebruikers jaarlijks van e-mail adres verandert (Rizzi, 2002). Voor eenmalige studies die kunnen beschikken over een recente database is dat wellicht geen onoverkomelijk probleem, maar bij longitudinale studies vormt dit wel een serieus probleem.

Het is vooralsnog niet mogelijk om de computers van willekeurige internetgebruikers direct te benaderen volgens RDD-achtige technieken, zoals ook zal blijken in paragraaf 3.4.

Spam

Misschien wel het grootste probleem van e-mail heeft echter te maken met *spam*, ongevraagde e-mail. Sinds de introductie van e-mail is in sneltreinvaart een enorme "*spam*"-industrie ontstaan die het veel moeilijker maakt om mensen ongevraagd via de e-mail te benaderen. Die "vermoeidheid" hebben we ook al kunnen constateren bij telefoon gesprekken, waarbij door de sterke groei van telemarketing personen ook minder geneigd zijn mee te werken aan een onderzoek. Bij e-mail is dat veel sneller gegaan, door de extreem lage kosten en de oncontroleerbaarheid van het verzenden van e-mail berichten is *spam* zeer sterk gegroeid en overheersend aanwezig in de e-mail inbox van veel argeloze internetgebruikers. Dientengevolge wordt ongevraagde e-mail van een onbekende al snel als *spam* aangemerkt, en dus ongelezen in de virtuele prullenbak gegooid. Onderzoek bevestigt dat respondenten uitnodigingen op de e-mail voor een onderzoek vaak beschouwen als *spam* (Topp & Pawloski, 2002). Het moet echter nogmaals opgemerkt worden dat het *spam* probleem vooral parten speelt als de afzender niet bekend is bij de ontvanger. In dat geval lijkt de geschiktheid van e-mail als communicatiekanaal voor het *eerste* contact voor verzoek om deelname aan een enquête dus ongeschikt.

Als de ontvanger van een e-mail echter wel bekend is met de afzender, is de kans op een positieve respons aanmerkelijk groter. Bij onderzoeken met een kleinere, specifieke populatie is het dus van belang om de e-mail te versturen via een organisatie die bekend is bij alle individuen. Zo zal een e-mail verzoek om deelname aan een tevredenheidsonderzoek onder studenten van de Universiteit Utrecht waarbij de naam "Universiteit Utrecht" vermeld wordt in het afzender- of onderwerpveld op een hogere respons kunnen rekenen dan wanneer de naam "Universiteit Utrecht" pas in de *body* van de e-mail vermeld wordt.

Bij onderzoeken met een grotere en meer algemene populatie is een dergelijke aanpak echter niet mogelijk. Om het toch mogelijk te maken respondenten via de e-mail aan te schrijven voor een verzoek aan deelname, zonder direct met *spam* geassocieerd te worden, moet van tevoren toestemming gevraagd worden aan de potentiële respondenten. Een dergelijke "opt-in" strategie, waarbij via diverse kanalen respondenten geworven worden om aan toekomstige onderzoeken medewerking te verlenen, heeft tot de vorming van *internet panels* geleid, welke in paragraaf 3.3 uitgebreid behandeld worden.

Andere methodes om respondenten te benaderen via websites, en dus een vorm van *convenience sampling* toe te passen, zijn reeds besproken in het overzicht van niet willekeurig getrokken steekproeven in paragraaf 2.5.1. Daar bleek reeds dat deze methodes de minst representatieve steekproeven opleveren, in andere woorden steekproeven waarbij een sterk *bias* aanwezig is.

Als laatste is het ook mogelijk om bij het trekken van de steekproef gebruik te maken van traditionele kanalen zoals de telefoon (RDD) of adresbestanden. Het grote voordeel hiervan is dat beproefde methoden gebruikt worden om een willekeurige steekproef te trekken. Een dergelijke *mixed-mode* methode van steekproef trekken kan de voordelen van online enquêteren echter alleen tot zijn recht doen komen indien meerdere malen gebruik gemaakt kan worden van een dergelijke *offline* steekproeftrekking. Wederom komen we dan uit bij *internet panels*.

3.1.3 Non-respons bias

Een andere bron die kan leiden tot onjuiste resultaten van een onderzoek, is het non-respons probleem. Een vertekend beeld van de resultaten ontstaat overigens pas als de non-respons onevenwichtig, ofwel niet random, verdeeld is over alle verschillende respondenten. Dit is echter meestal het geval.

Een andere voorwaarde voor onjuiste weergave van resultaten is het bestaan van een relatie tussen de verschillende typen respondenten en non-respondenten, en de variabelen aan de hand waarvan onderzoeksresultaten bekend worden gemaakt. Zo zal de uitkomst van een onderzoek naar het rechts- of linkshandig zijn van mensen niet wezenlijk verschillen als blijkt dat hoger opgeleide groepen fors oververtegenwoordigd zijn. Maar bij een onderzoek naar het al dan niet beschikken over een internetverbinding zal dit juist wel tot niet-representatieve resultaten leiden.

Een grof onderscheid is te maken naar zogenaamde eenheid non-respons (*unit non-respons*) en item non-respons. Bij eenheid non-respons gaat het om respondenten die wel tot het steekproef kader behoren, maar van wie om diverse mogelijke redenen geen gegevens bekend zijn, oftewel die de (online) enquête niet hebben ingevuld. Bij item non-respons gaat het om enkele vragen die niet beantwoord zijn. Aanbevelingen om item non-respons te voorkomen komen in de volgende paragraaf over meetfouten aan bod, in deze paragraaf staat de eenheid non-respons centraal.

Redenen voor non-respons

Er zijn vele mogelijke redenen voor non-respons. Crawford (2001) hield in een studie naar non-response in web enquêtes een uitgebreid na-onderzoek onder de non-respondenten en wist zo de redenen voor non respons naar liefst 12 verschillende categorieën te onderscheiden. Het voert te ver om deze allemaal te behandelen, maar op basis van deze en andere publicaties (Dillman & Bowker, 2001 en Manfreda & Vehovar, 2002) zijn op hoofdlijnen de volgende algemeen geldende redenen te onderscheiden.

Respondenten horen (toch) niet bij de populatie en/of steekproef

Indien de te onderzoeken populatie een specifiek karakter heeft, moet reeds in de eerste vragen van de enquête vast te komen staan of de respondent tot de te onderzoeken populatie behoort. Zo zal de eerste vraag in een enquête onder studenten behoren te zijn "studeert u?" of een vraag met een dergelijke strekking.

Deze vorm van non-respons vormt in principe geen gevaar voor de kwaliteit van de steekproef, omdat onterechte geselecteerde respondenten er juist uitgefilterd worden. Het enige mogelijke nadeel is een afname van de werkelijke steekproefgrootte.

Respondenten kunnen niet bereikt worden

Ging het hier in het geval van telefonische enquêtes vooral om telefoonnummers waarbij niet opgenomen wordt, bij online enquêtes is deze onbereikbaarheid iets anders van aard. Het enige gangbare contact bij online onderzoek vindt plaats via de e-mail. Bij e-mail worden doorgaans onbestelbare emails teruggezonden naar de verzender/onderzoeker. Zo kan de onderzoeker vaak reeds binnen enkele minuten een overzicht hebben van e-mail adressen die niet "kloppen".

Toch wil andersom het niet ontvangen van een dergelijke foutmelding niet zeggen dat de respondent daadwerkelijk de e-mail ontvangen heeft. Is een telefoonnummer en adres nog redelijk onveranderlijk, bij een e-mail adres is het tegendeel waar. Zo is het niet ongebruikelijk om meerdere adressen te gebruiken, maar ook regelmatig van adres te

veranderen. Het is daarom nooit uit te sluiten dat de e-mail ongelezen ergens op een "mailserver" in het spreekwoordelijke digitale universum blijft staan.

Respondenten ondervinden technische problemen

Deze non-respons kan twee oorzaken hebben. Enerzijds kunnen er daadwerkelijk technische problemen optreden als gevolg van een fout ontwerp van de enquête. Een van de kenmerkende verschillen tussen online enquêtes en papieren enquêtes, is dat bij de online enquêtes dat, wat de ontwerper van de enquête ziet op zijn scherm, niet overeen hoeft te komen met wat respondenten zien op hun scherm (Dillman & Bowker, 2001). Zo kan een enquête op computers met een bepaalde dikwijls wat oudere configuratie voor problemen zorgen. Het is aan de ontwerper van de enquête om rekening te houden met diverse configuraties en internetaansluitingen en de enquête hierop aan te passen.

Anderzijds is het ook mogelijk dat de respondent over onvoldoende kennis beschikt om de enquête te openen of in te vullen. Dillman & Bowker (2001) noemen bijvoorbeeld het niet om kunnen gaan met verschillende antwoordmethoden zoals *radio buttons*, *html boxes* en *drop-down menus*.

In de volgende paragraaf wordt uitgebreider ingegaan op het ontwerp van een enquête en de gevolgen voor eventuele non-respons.

Respondenten kunnen zich ontvangst enquête(/verzoek) niet herinneren

Volgens de vergelijkende studie naar online en offline enquêtes van Crawford e.a. (2002) is veruit de meest genoemde reden voor non-respons bij online enquêtes dat mensen zeggen zich niet te kunnen herinneren de uitnodiging te hebben ontvangen, ondanks de juiste contact informatie. Opvallend is hierbij dat deze reden alleen eruit springt bij online enquêtes, bij enquêtes verstuurd via de normale post was "geen tijd" de meest genoemde reden voor non-respons. Waarschijnlijk maakt het ontvangen van een e-mail een minder blijvende indruk op potentiële respondenten dan het ontvangen van een papieren brief.

Respondenten willen niet meewerken aan enquête

Zoals ook uit de cijfers uit de studie van Crawford e.a. (2001) blijkt, bestaan er bij deze vorm van non-respons niet veel wezenlijke verschillen tussen traditioneel onderzoek en online onderzoek. Als een respondent "geen zin", "geen tijd" of "geen interesse" heeft, heeft de onderzoeker zich daarbij neer te leggen. Voordat het zover komt, kan de onderzoeker er echter wel alles aan doen om de interesse van de potentiële respondent te wekken, bijvoorbeeld door nauwkeurig aandacht te besteden aan de tekst op de uitnodiging. In hoeverre deze en andere factoren van invloed zijn op de responspercentages in zijn algemeen, komt in de volgende paragraaf over meetfouten aan bod.

Maar eerst besluiten we deze paragraaf met een tweetal aandachtspunten die van belang zijn bij het beperken van non-respons.

Beperken van non respons

Contacten

Vrijwel overal in de literatuur omtrent responspercentages, ongeacht of het om traditioneel of online onderzoek gaat, wordt het belang onderstreept van meervoudige contacten met de respondent, zowel vóór (*pre-notification*) als ná (*non-response follow-up*) het versturen van de enquête (Schonlau e.a. 2001, Dillman 2000, Manfreda & Vehovar 2002).

Een uitgebreide literatuurstudie van Cook e.a. (2000) naar de responspercentages van e-mail en web enquêtes bevestigt deze bevindingen. De onderzoekers namen 49 studies onder de loep die voldoende informatie over responspercentages en de gebruikte methodologie

bevatten over in totaal 68 verschillende enquêtes. Het gemiddelde responsepercentage bedroeg 39,6%, met een standaarddeviatie van 19,6%. Een regressie analyse werd vervolgens uitgevoerd om verschillende factoren te herkennen die van invloed zijn op de responspercentages van e-mail en web enquêtes. Geconcludeerd werd dat het *aantal contacten*, *gepersonaliseerde contacten* en *contact vooraf* de drie belangrijkste factoren waren die positief correleerden met responspercentages. Het aantal contacten heeft betrekking op de aantal keren dat een herinnering gestuurd wordt naar respondenten die nog niet gereageerd hebben. Met *gepersonaliseerde contacten* wordt hier bedoeld dat in de contacten (brieven) de respondent bij naam genoemd wordt. *Contact vooraf (pre-notification)* tot slot, betekent dat de respondent middels een kort bericht op de hoogte wordt gebracht dat hij/zij binnenkort benaderd zal worden voor een belangrijk onderzoek. Behalve Cook e.a. concluderen ook Manfrede & Vehovar (2002) dat contact vooraf leidt tot zowel een lagere unit non-respons als item non-respons.

In traditioneel onderzoek betekende elk extra contact met respondenten die (nog) niet gereageerd hadden een extra aanslag op het budget van het onderzoek, zowel door administratieve kosten als marginale kosten per contact (aan callcenter of papier & postzegels). Bij online onderzoek, waarbij contact doorgaans via de e-mail plaatsvindt, vallen deze extra kosten echter aanzienlijk lager uit. En zeker als gebruik gemaakt wordt van gespecialiseerde enquête-software kunnen alle uitgaande herinneringsberichten verstuurd worden zonder dat daar een persoon aan te pas hoeft te komen.

Incentives

Door een zogenaamde *incentive* aan te bieden, kan de onderzoeker "iets terug doen" voor de respondent als dank voor het meewerken aan een enquête. De werkelijke reden heeft echter meestal meer te maken met het aantrekken van meer respondenten. In het geval van online enquêtes kan geprobeerd worden om ook een online cadeau te geven, zoals een zogenaamde *electronic gift certificate* te besteden bij een internet winkelier, maar meer gebruikelijk in Nederlandse enquêtes is deelname aan een soort loterij waarin prijzen te winnen zijn.

Hoewel het bieden van een *incentive* op het eerste gezicht een logische methode lijkt om de respons te verhogen, is in diverse onderzoeken (Cook e.a., 2000, Manfrede & Vehovar, 2002 en Dillman 2000) gebleken dat een positieve relatie tussen het bieden van *incentives* en responspercentages niet zo logisch is. Een positieve invloed van *incentives* op responspercentages kon niet aangetoond worden.

En als de *incentive* niet direct aangeboden wordt maar pas na het volledige invullen van de enquête, kan zelfs een lager responspercentage het gevolg zijn. In dat geval kan de potentiële respondent namelijk de interactie als een soort economische transactie gaan beschouwen, waar zonder *incentive* nog sprake zou zijn van een sociale interactie van *goodwill*. Vervolgens kan de potentiële respondent het idee hebben dat hij een moreel legitieme keuze heeft om al dan niet mee te werken aan de enquête, en omdat de *incentive* in geldelijke termen veelal klein is kan hij besluiten dat de *incentive* de moeite niet waard is om te reageren (Schonlau e.a., 2001).

Tot slot heeft een gebruikersvriendelijke opzet en ontwerp van de enquête een positieve invloed op het beperken van zowel de unit non-respons maar vooral ook de item non-respons. De volgende paragraaf gaat hier verder op in.

3.1.4 Meetfouten

Meetfouten ontstaan als het gegeven antwoord in een enquête verschilt van het "werkelijke", "echte" antwoord. Zo kunnen respondenten bijvoorbeeld een vraag niet goed begrijpen, en vervolgens een antwoord geven dat niet overeenkomt met het werkelijke antwoord op de vraag zoals de onderzoeker het bedoeld heeft. Meetfouten zijn nooit helemaal uit te sluiten, maar worden wel beïnvloed door de methode van gegevensverzameling. Met betrekking tot online onderzoek speelt met name het ontwerp van de enquête een grote rol bij het beperken van meetfouten, deze paragraaf gaat met name daar verder op in.

Bij de eerste enquêtes die verstuurd werden over het internet, ging het vooral om e-mail formulieren. Beperkt tot de text-only mogelijkheden van destijds, moest de respondent een "x" plaatsen tussen twee haakjes [x] indien het antwoord van toepassing was. Het verwerken van dergelijke e-mail formulieren met ook open vragen in een database of spreadsheet, vereiste relatief veel tijd en kennis van computersystemen. Bovendien werden voordelen die het gebruik van computers konden bieden in het ontwerp van een enquête, op deze manier nauwelijks gebruikt (Topp & Pawloski, 2002).

Dat veranderde met de komst van het world wide web, waarbij web formulieren gemaakt konden worden. Het grootste voordeel van webformulieren is dat bij het ontwerpen de maker grotere vrijheid en een scala aan nieuwe mogelijkheden heeft voor het presenteren van enquêtes. Toch ligt juist daar ook een gevaar op de loer, want het grote aantal design mogelijkheden kan ook resulteren in het maken van de verkeerde keuzes. Keuzes waarbij de functionaliteit en de respondent niet meer centraal staan. Voor betrouwbare resultaten, en voorkomen van partiele non-respons, is het dus van belang dat veel zorg wordt besteed aan een functioneel ontwerp van de enquête.

Aanwijzingen

De "goeroe" op het gebied van goede doordachte en wetenschappelijk verantwoord ontworpen enquêtes is Don. A. Dillman, die werkzaam is als professor op de faculteit van sociale wetenschappen op Washington State University. Hij heeft talloze publicaties op het gebied van enquêteontwerp op zijn naam heeft staan, waaronder twee standaardwerken (1978, 2000) die in vrijwel geen literatuurlijst over dit onderwerp ontbreken. Zijn boek "Mail and Internet Surveys, the Tailored Design Method" (2000) legt in 400 pagina's uit hoe een goede enquête er uit moet zien.

Het onderzoek naar het ontwerp en de implementatie van online enquêtes leent zich echter minder goed voor een standaardtheorie van regels en concepten (Schonlau e.a., 2001). Omdat er een divers spectrum aan verschillende factoren van belang zijn volgt hier een opsomming die de belangrijkste aandachtspunten behandelt met betrekking tot het ontwerp van een webenquête.

Aantal vragen per scherm

Een belangrijke keuze die gemaakt moet worden bij het opzetten van een online enquête, is die tussen het publiceren van de enquête op één pagina (waarbij de respondent naar beneden moet *scrollen* om verdere vragen te beantwoorden) en het presenteren van één of enkele vragen per pagina, waarna de respondent op een knop moet drukken om verder te gaan ("*volgende vraag*").

Dillman (2000) geeft de voorkeur aan enquêtes op één pagina waarbij van het begin tot het eind van de enquête *ge-scrolled* kan worden. Hij stelt dat deze manier de dominante vorm geworden is bij het ontwerp van de meeste web enquêtes, en haalt daarbij een bron aan

(Bowker, 1999 in Dillman, 2000) die aantoont dat 90% van de onderzochte web enquêtes een één pagina scrolling constructie hadden, in plaats van een *multi-screen* constructie.

De mening van Schonlau e.a. (2001) staat hier duidelijk lijnrecht tegenover, zij adviseren om slechts enkele vragen per scherm weer te geven. Volgens hen kan uitgebreid scrollen tot frustratie leiden bij respondenten, en is ook uit onderzoek gebleken dat bij een *single screen* aanpak de *drop-out rate*, dat wil zeggen het tussentijdse afhaken van respondenten, groter is. De werkwijze van gerenomeerde onderzoeksbureaus, die grote internetpanels beheren, geven de methode met meerdere schermen, dus zonder *scrollen*, gelijk. Bij vrijwel al de enquêtes zien de respondenten slechts enkele vragen per scherm.

Een bijkomend voordeel van een dergelijke multi-screen aanpak is dat antwoorden makkelijk tussentijds opgeslagen kunnen worden. In het geval dat de respondent de enquête plotseling verlaat, gaat zo niet onnodig informatie verloren.

Antwoordmogelijkheden

Over zeer veel verschillende, soms op het eerste oog insignificante details, is inmiddels onderzoek gedaan. Zo blijkt uit onderzoek van Heerwegh e.a. (2000) dat zogenaamde dropdown boxes moeilijker te gebruiken zijn dan radiobuttons. Zij concluderen dat vooral als niet-ervaren internet gebruikers meedoen, radiobuttons de voorkeur verdienen.

Forceren van antwoorden

Bij online enquêtes bestaat de mogelijkheid om antwoorden te forceren, door respondenten te dwingen een antwoord aan te kruisen op een meerkeuze vraag voordat ze door kunnen gaan naar de volgende vraag. Als in een dergelijk geval geen “geen antwoord / geen mening”-categorie beschikbaar is, kan gesproken worden over het forceren van antwoorden. Macroy e.a. (2002) concludeerden in hun studie naar verschillen tussen webenquêtes en papieren enquêtes dat het forceren van antwoorden respondenten niet afschrikt van het afmaken van een enquête. Integendeel, de conclusie wees erop dat het forceren kon leiden tot een meer complete dataset om te analyseren.

Toch raden zowel Dillman (2000) als Schonlau e.a. (2001) de ontwerper van online enquêtes af om gebruik te maken van de mogelijkheid tot het forceren van antwoorden. Terecht merken ze op het de respondent vrij staat om geen antwoord in te vullen, en het forceren van antwoorden vrij makkelijk kan leiden tot irritatie en frustratie bij de respondent, mogelijk resulterend in het voortijdig afbreken van de enquête. De enige uitzondering op deze regel geldt voor zogenaamde *screening* vragen, die beantwoord moeten worden vóór de aanvang van de werkelijke enquête.

Indicatie voortgang

Bij een papieren enquête kan de respondent makkelijk zien hoeveel vragen hij nog moet beantwoorden. Dit is bij een online enquête alleen mogelijk als gebruik wordt gemaakt van een scherm waardoor de respondent door te scrollen wel een idee kan krijgen over de voortgang.

De meeste online enquêtes maken echter gebruik van meerdere schermen, en dus kan de respondent de indruk krijgen een schijnbaar eindeloze serie van vragen voorgeschoteld te krijgen. Zo blijkt uit waarnemingen van verschillende onderzoekers (Dillman, 2000) dat bij online enquêtes regelmatig respondenten een paar vragen voor het eind stoppen met de enquête. Een goede indicator kan deze vorm van non-respons beperken. Bij voorkeur minimaal drie keer (na elke kwart afgemaakte enquête) moet de respondent op de hoogte gebracht worden van zijn voortgang.

Hiermee samenhangend kan ook vermeld worden dat enquêtes nooit te lang mogen duren. In het algemeen wordt een richtlijn aangehouden van maximaal 15 a 20 minuten.

Beveiliging

Het internet is een wereldwijd netwerk, en enquêtes die online staan, zijn dus ook per definitie toegankelijk voor het hele internetpubliek. Om oneigenlijk gebruik uit te sluiten, is het noodzakelijk om met een unieke identificatie code en password te werken. Eventueel kunnen deze in het internetadres van locatie van de enquête verwerkt worden, waardoor extra inspanning van de kant van de respondenten beperkt kan blijven.

Foutmeldingen

Eventuele foutmeldingen dienen zo specifiek mogelijk te zijn. Tenminste moet duidelijk zijn bij welke vraag een eventuele fout opgetreden is, en een duidelijke omschrijving van de fout is ook aan te bevelen ("U bent vergeten vraag 21 in te vullen").

Gebruik media mogelijkheden

Door gebruik te maken van de mogelijkheden die het world wide web biedt bij het presenteren van informatie, kan de online enquête gebruiksvriendelijk gepresenteerd worden. Zo kunnen bijvoorbeeld help-links verwijzen naar een pop-up window met extra informatie over de desbetreffende vraag, zonder dat de enquête langer of onoverzichtelijker hoeft te worden. En door het gebruik van kleuren kan de structuur van een online enquête makkelijker begrepen worden door een respondent. Het gebruik van kleuren en afbeeldingen moet echter wel functioneel blijven, de enquête is er immers niet voor gemaakt om op te vallen. Dillman (1998) heeft aangetoond dat een "kale" enquête resulteert in een iets hoger respons percentage dan een "fancy" vormgegeven enquête.

Gebruik automatisering mogelijkheden

De mogelijkheden die diverse internet programmeertalen inmiddels bieden bij het ontwikkelen van interactieve internet pagina's zijn eindeloos. Toch moet met zorg (en mate) omgegaan worden met deze mogelijkheden, want een te intensief gebruik hiervan kan enerzijds tot aanzienlijk hogere setup-kosten leiden en anderzijds tot een hogere non-respons omdat bepaalde (oudere) configuraties van computers van respondenten mogelijkwijs geen toegang kunnen krijgen tot de enquête. Maar indien met zorg opgebouwd, kunnen wel zeer handige automatiseringsfuncties geïmplementeerd worden die ten goede komen aan de kwaliteit van de enquête zonder dat dat tot problemen hoeft te leiden.

Door de software de zogenaamde *skip* patronen te laten beheren, worden fouten voorkomen en wordt de enquête makkelijk in te vullen vanuit het perspectief van de respondent.

Een vrij makkelijk te implementeren mogelijkheid die tot betere gegevensverzameling kan leiden, is het random "husselen" van de volgorde van antwoordcategorieën met een nominale schaal. Zo krijgt elke respondent een andere (random) volgorde van antwoorden op zijn scherm te zien, waardoor elke antwoordcategorie in principe een gelijke uitgangspositie heeft.

Ook is het mogelijk antwoorden te controleren op juistheid tijdens het afnemen van de enquête. Zo kan de respondent bijvoorbeeld een waarschuwing op zijn scherm krijgen als hij/zij bij de geboortedatum een onmogelijke datum (31 februari) invult. Dergelijke controle mechanismen moeten echter gebruiksvriendelijk en niet forcerend van opzet zijn, want uiteindelijk moet de respondent de eindverantwoordelijkheid gegeven worden over zijn eigen antwoorden.

Verder kan gewezen worden op de extra informatie die de computer over het proces van invullen kan genereren. Zo kan bijvoorbeeld berekend en genoteerd worden hoe lang de respondent over een enquête doet. Vervolgens kan ervoor gekozen worden om respondenten die de enquête binnen een onwaarschijnlijk snelle tijd afronden, niet verder mee te nemen in de gegevensverwerking. Het Nederlandse onderzoeksbureau Multiscope (2003) gebruikt

deze techniek om respondenten die alleen "voor de prijs" snel door enquêtes heen klikken, uit de steekproef te filteren.

Aangezien het gebruik van internet in allerlei processen in de samenleving nog zo kort geleden begonnen is, moeten zowel de ontwerpers van web applicaties als internetgebruikers gewend raken aan dit nieuwe medium. Het is nu nog te vroeg om uitspraken te kunnen over hoe de cognitieve kwaliteit van online enquêtes zich verhoudt tot die van traditionele enquêtes. Wel heeft onderzoek volgens Topp & Pawloski (2002) uitgewezen dat cognitieve verschillen tussen online enquêtes en traditionele enquêtes alleen aangetoond kunnen worden op longitudinale basis. Dit zou kunnen suggereren dat respondenten meer en meer gewend raken aan de technologie en het world wide web.

De schrijvers merken dan ook op (2002, p. 176):

"It is worth remembering that many of us are immigrants to this technology (internet), while our children are native to it."

3.2 Weegmethoden

Zolang een willekeurige steekproef getrokken wordt uit het hele steekproefkader, en er geen steekproeffouten ontstaan door enige vorm van non-respons is het wegen van via enquête verzamelde gegevens niet nodig. Ieder individu in de populatie heeft immers een gelijke kans om in de steekproef getrokken te worden.

Dit is echter zelden het geval. Met name als de te onderzoeken populatie een algemeen karakter heeft, zoals alle inwoners van een bepaald land, kan zelden een echt “puur” willekeurige steekproef getrokken worden. Doordat bepaalde groepen mensen met kenmerken die niet random verdeeld zijn, over- of ondervertegenwoordigd zijn, kunnen *biasen* optreden in de resultaten. Een van de manieren om voor deze *biasen* na het afnemen van enquêtes te corrigeren, is het wegen van de gegevens. Al snel na de introductie van steekproef onderzoek werd weging toegepast voor het correct weergeven van resultaten. Wegingprocedures worden gebruikt voor de correctie van unit non-respons. Zij kennen extra gewicht toe aan cases in een stratum dat onder-gerepresenteerd is. Het eindresultaat is dat alle soorten cases in de juiste verhouding voorkomen in de steekproef (Gouw & Groenland, 2000).

In deze paragraaf zullen de twee belangrijkste en meest gehanteerde weegmethodes besproken worden zoals die gebruikt worden bij het wegen van online enquêteresultaten. De eerste, gestratificeerd wegen, behelst de klassieke manier van wegen zoals die al decennia toegepast wordt op allerlei steekproeven. Hierbinnen worden verschillende methoden onderscheiden, die gemeen hebben dat zij werken met het toekennen van gewichten aan verschillende deelgroepen, zogenaamde strata.

De tweede methode, het zogenaamde *propensity weighting*, is nieuwer. Zij vindt haar oorsprong in de medische wetenschappen, maar is redelijk recent geïntroduceerd bij het corrigeren van gegevens verkregen door online enquêtes. De vraag in hoeverre propensity weighting als oplossing kan dienen bij het betrouwbaar wegen van enquêteresultaten verkregen door middel van online enquêtes, verdient hierbij bijzondere aandacht.

3.2.1 Gestratificeerd wegen

Al snel na de introductie van steekproefonderzoek, beseften wiskundigen dat de kennis die reeds voorhanden was over een populatie, op de een of andere manier moest worden benut bij het schatten van kenmerken, die representatief voor diezelfde populatie geacht werden te zijn.

Neyman liet als eerste het nut zien van het gebruik van reeds bekende informatie, door een wiskundig onderbouwde (pre-)stratificatiemethode te publiceren (Sikkel, 2000). Het splitsen van een steekproef in verschillende strata kon zowel vóór als ná het trekken van de steekproef plaatsvinden. Indien het stratificeren vóór de steekproeftrekking plaatsvindt, is eigenlijk geen sprake van het *wegen* van enquêteresultaten, maar voor de volledigheid wordt deze methode toch kort besproken in deze paragraaf. In het laatste geval wordt van poststratificatie gesproken, wat de meest eenvoudige en gebruikelijke wegingmethode is. Andere voorbeelden van vergelijkbare wegingprocedures die hier ook aan bod komen zijn additief (lineair) wegen en multiplicatief wegen. Aangezien de vraag op welke variabelen gewogen dient te worden voor de verschillende genoemde weegmethoden niet wezenlijk verschilt, wordt dit aan het eind van deze paragraaf besproken.

Pre-stratificatie

Bij deze oude stratificatiemethode wordt voorafgaande aan het trekken van de steekproef, het steekproefkader opgesplitst in verschillende deelsteekproeven. Op deze wijze kan uiteindelijk een steekproef verkregen worden, waarbij de verschillende deelpopulaties proportioneel verdeeld zijn. Hiertoe moet het wel mogelijk zijn van te voren te kunnen bepalen welke aantallen respondenten uit strata waarvan de populatieomvang bekend zijn, aan het onderzoek deelnemen.

Als praktijkvoorbeeld hiervan met betrekking tot online onderzoek, kan de oververtegenwoordiging van breedband internet gebruikers in internet panels dienen. Online onderzoeksbureau Multiscope (2003), heeft een internetpanel in beheer waarvan verschillende kenmerken van de deelnemers bekend van zijn. Het blijkt dat relatief veel breedband internet gebruikers deelnemen aan het panel, veel meer dan proportioneel verwacht zou worden op basis van landelijke internetcijfers uitgesplitst naar type verbinding. Door vóór het trekken van de steekproef, de steekproef te stratificeren over verschillende strata (lees: typen internetverbinding), wordt de representativiteit van de totale steekproef groter.

Een ander voordeel is bovendien dat doorgaans de variantie van een te schatten variabele afneemt ten opzichte van een aselechte, niet gestratificeerde steekproef. Dit is mogelijk gemaakt door de informatie die over de omvang van verschillende strata in de populatie bekend is. Opgemerkt zij wel dat stratificatie ook tot variantieverhoging en vertekening van de te schatten variabelen kan leiden, indien de gewichten niet goed worden berekend (Sikkel, 2000)

Poststratificatie

Vaak kan van te voren niet bepaald worden in welke strata respondenten vallen, simpelweg omdat (juist) daarvoor de respondenten eerst geënquêteerd moeten worden. Om informatie over omvang van de populatie verdeeld over verschillende strata toch te kunnen gebruiken, ging men daarom over tot poststratificatie. Hierbij worden achteraf gewichten toegekend die zodanig zijn, dat de gewogen verdeling over de strata correspondeert met de populatieverdeling. Alle steekproefelementen binnen een stratum a word een gewicht w_a toegekend, die als volgt berekend kan worden (waarbij N is populatieomvang en n is steekproefomvang).

$$w_a = \frac{n * N_a}{n_a * N}$$

Kon aanvankelijk nog maar op een beperkt aantal hulpvariabelen gewogen worden, sinds de computers hun intrede hebben gedaan is het ambitieniveau van de onderzoekers op het gebied van de statistiek toegenomen. Populatiegegevens kunnen met behulp van meerdere hulpvariabelen worden uitgesplitst in meerdimensionale tabellen met talloze cellen. Dit brengt echter ook een beperking aan het licht; naarmate op meer hulpvariabelen gewogen wordt neemt de kans op lege of nauwelijks gevulde cellen exponentieel toe. In dat geval kan ofwel het gewicht niet bepaald worden (delen door 0 kan immers niet), ofwel bestaat de kans dat er een veel te groot gewicht wordt toegekend aan de antwoorden van één of enkele respondenten.

Om toch op meerdere hulpvariabelen te kunnen wegen zijn nieuwe benaderingen ontwikkeld waarvan de belangrijkste, additief en multiplicatief wegen, hieronder besproken worden.

Additief en multiplicatief wegen

Aan de basis van deze nieuwe benaderingen, die echter nog wel gebaseerd zijn op stratificatie als uitgangspunt, staat de vooronderstelling dat alle cellen binnen een tabel niet per se gereproduceerd moeten kunnen worden, het is voldoende als de randtotalen kloppen. De populatieverdelingen van de hulpvariabelen moeten dus wel overeenkomen met de verdeling in de steekproef, maar binnen cellen van de meerdimensionale tabel hoeft dit niet noodzakelijk. Het idee is dat voor alle hulpvariabelen voor elk afzonderlijk stratum een soort van dummy-variabele, ofwel gewichtscomponent, bepaald wordt. De gewichtscomponenten zijn vergelijkbaar met de waarden van de coëfficiënten in een regressievergelijking. Met de gewichtscomponenten van strata die van toepassing zijn op een bepaalde cel kan vervolgens een berekening worden uitgevoerd om te komen tot het gewicht van de individuele cel (Bethlehem, 1997).

Alvorens deze berekening uit te kunnen voeren, om zo de individuele gewichten van alle steekproef respondenten afzonderlijk te bepalen, is het nodig extra randvoorwaarden op te stellen. In de jaren 70 werden twee typen randvoorwaarden geformuleerd, die gebaseerd zijn op de algemene theorie van regressie-schattingen. De randvoorwaarde die stelt dat de *som* van de gewichtscomponenten gelijk moet zijn aan het uiteindelijke gewicht van individuele cellen, wordt gehanteerd bij additief (of lineair) wegen. Additief wegen heeft als voordeel dat de methode, in tegenstelling tot die van poststratificatie, ook te gebruiken is bij numerieke variabelen. Nadeel van de methode is dat de gewichten negatief kunnen worden. Geldt als randvoorwaarde dat het *produkt* van de gewichtscomponenten gelijk moet zijn aan het uiteindelijke gewicht van individuele cellen, dan is er sprake van multiplicatief wegen. Deze methode wordt het meest toegepast omdat het de natuurlijke uitbreiding van poststratificatie is. Ze is uitsluitend toepasbaar op categorische variabelen, maar heeft het voordeel dat alle individuele gewichten positief zijn.

Weegvariabelen

De hulpvariabelen waarop gewogen moet worden, dienen liefst een zo hoog mogelijk verklarend vermogen te hebben, en dus een duidelijke samenhang te hebben met de doelvariabelen uit het onderzoek. Aan de andere kant moeten van deze hulpvariabelen ook exacte populatiegegevens beschikbaar zijn. De variabelen waar men dan meestal op uitkomt hebben veelal een sociaal-demografisch karakter.

Zoals geconstateerd in paragraaf 3.1.1 is de online populatie anders samengesteld dan de populatie in het algemeen. Zo zijn laag opgeleide mensen, huishoudens met lage inkomens en ouderen bijvoorbeeld vaak ondervertegenwoordigd. Maar ondanks dat de percentages laag zijn, blijken ook bij deze groepen mensen grote absolute aantallen online te zijn. Dus, met de juiste steekproefgrootte, is het mogelijk om online data te wegen om representatief te zijn op sociaal-demografisch variabelen zoals leeftijd, geslacht, regio, opleiding, beroep, inkomen, etniciteit, woonomgeving en gezinssamenstelling.

Als gezocht wordt naar nieuwe verklarende variabelen binnen verschillende groepen respondenten van online enquêtes, valt op dat vooral daar waar sprake is van een *self-selecting* element in de werving van respondenten, bepaalde groepen zoals heavy-users, gebruikers met snelle verbindingen en kopers op het internet zijn oververtegenwoordigd (Multiscope, 2003). In dit kader lijkt het belangrijk om de juiste hulpvariabelen waarop gewogen kan worden te kiezen. Andere mogelijke variabelen die verschillen binnen de groep internetgebruikers kunnen betrekking hebben op het eerste jaar van internetgebruik, e-mail gebruik, locatie van internettoegang (werk/thuis) en algemene computer oriëntatie (Vehovar & Manfreda, 1999). De beperkende factor hierbij is de vooralsnog gebrekkige informatie die

beschikbaar is over enerzijds de kenmerken van respondenten en non-respondenten van online enquêtes, en anderzijds de populatiestrata, die bovendien beiden onderhevig zijn aan de snelle veranderingen die kenmerkend zijn voor de ontwikkeling van internet.

3.2.2 'Propensity weighting'

Achtergrond

Een onderzoeker kan een zelfde enquête aan twee groepen voorleggen, waarbij bij elke groep een andere methode van enquêteren gehanteerd wordt. Als vervolgens significante verschillen tussen de verschillende groepen waargenomen worden, kunnen daar twee oorzaken aan ten grondslag liggen. Ofwel er bestaan daadwerkelijk verschillen tussen de kenmerken van beide groepen, ofwel verschillen zijn het gevolg van de verschillende methode van enquêteren.

Vergelijkende statistische methoden moeten in dit geval duidelijkheid bieden, waarbij gesproken kan worden van zogenaamd "*multi-method*" onderzoek. Onderscheid kan gemaakt worden naar het experimentele of observerende karakter van het onderzoek, en een onderzoeksvraag gerelateerd aan kenmerken binnen of tussen de groepen respondenten (Miller & Panjikanan, 2001).

Op het gebied van online onderzoek, gaat het doorgaans om de vraag of enquêteresultaten van online respondenten groepen vergelijkbaar zijn met die van traditionele respondenten groepen. Wanneer er van alle antwoorden gelijke verdelingen waargenomen worden, kan gesteld worden dat traditionele en online enquête methoden vergelijkbaar zijn. Deze vergelijking kan plaatsvinden met ongewogen gegevens, maar ook met gewogen gegevens.

In voorgaande paragraaf zijn reeds gangbare stratificatie weegmethoden besproken, maar deze methoden hebben vooral betrekking op het wegen van steekproefgegevens naar de populatie kenmerken. Een methode die meer geschikt is voor het vergelijkbaar maken van twee verschillende steekproeven, is *propensity weighting*.

Propensity weighting methode in een notedop

Als statistische methode in de medische wetenschap ontwikkeld door Rosenbaum (Danielsson, 2002), had zij als doel het direct vergelijkbaar maken van twee groepen (en steekproeven) die gescheiden zijn door middel van observationele variabele, zoals het al dan niet roken. In wezen worden de twee steekproeven vergelijkbaar gemaakt door tegelijkertijd te corrigeren voor alle variabelen die mogelijk van invloed zijn op de vergelijking. De logistische regressie-analyse die hiervoor gebruikt wordt, levert een schatting op van de kans dat een waarneming tot een van de beide groepen behoort, bijvoorbeeld de groep van online ondervraagden of telefonisch ondervraagden. Deze kansen kunnen worden uitgedrukt in de propensity scores (de Gouw en Groenland, 2000).

Herweging van een doelbestand vindt plaats door de propensity-score-verdeling van het doelbestand gelijk te maken aan de propensity-score-verdeling van het referentiebestand. Hierbij dient de propensity-score-verdeling van het referentiebestand te worden opgesplitst in 5 gelijke groepen, ofwel quintielen. De randwaarden die bij deze quintielen horen, worden overgebracht op het doelbestand, waarna de percentages van de ontstane groepen in het doelbestand bepaald kunnen worden. De laatste stap is het maken van een weegfactor voor elk quintiel, welke het percentage van het referentiebestand, gedeeld door het percentage van het doelbestand is.

Als verschillen tussen online en telefonische antwoorden veroorzaakt worden door verschillen tussen de groepen respondenten die onderdeel uitmaken van de variabelen waarop de weging toegepast is, dan zou het op propensity scores gewogen doelbestand resultaten moeten weergeven die overeenkomen met de resultaten van het bronbestand.

Een belangrijke vooronderstelling die vermeld moet worden, is die van de zogenaamde “*strong ignorability*” (Danielsson, 2002). Deze stelt dat de weging alleen correcte resultaten oplevert als geldt, dat de respondenten van beide groepen dezelfde verdeling van antwoorden zouden geven, wanneer zij dezelfde antwoorden geven op de variabelen waarop gecorrigeerd wordt.

Propensity weighting bij online onderzoek

Propensity weighting bij online onderzoek wordt vooral gebruikt voor de correctie van specifieke systematische non-respons die optreedt bij niet-willekeurig getrokken steekproeven online, waar met name bedoeld wordt op de zogenaamde self-selection bias. Deze bias is doorgaans niet te verklaren uit demografische variabelen, waardoor traditionele weegmethoden tekort schieten. Propensity weighting zou in dit geval uitkomst kunnen bieden, onder meer omdat het mogelijk is om veel verschillende en ook niet-demografische variabelen te wegen.

Hoewel deze theorie in de medische wetenschappen veel gebruikt en dus ook gedocumenteerd is, is de hoeveelheid literatuur die *propensity scoring* benadering bestudeert met betrekking tot de online-offline vergelijkende steekproef nog zeer beperkt. Daarnaast moet met zorg gekeken worden naar de bron van publicaties, aangezien online onderzoeksbureaus, die vooralsnog een stevige theoretische basis met betrekking tot hun online enquêtewerkzaamheden moeten ontberen, belang hebben bij een positief oordeel van deze relatief nieuwe methodiek.

Zo houdt Humprey Taylor (1999), president van een van de grootste internetpanels van Amerika (Harris Online Panel), een pleidooi voor het gebruik van *propensity weighting* bij online onderzoek, waar hij met name verwijst naar succesvol behaalde resultaten bij verkiezingspeilingen naar politieke voorkeur. Hij stelt dat de *propensity weighting* benadering succesvol kan zijn in het wegen van online enquêteresultaten naar gegevens van parallel gehouden telefonische enquêtes. Met name door de mogelijkheid om op vele variabelen te kunnen wegen zonder de foutmarge van de schattingen te verhogen, verdient deze methode de voorkeur boven traditionele weegmethoden, zo stelt Taylor.

Om het grote voordeel van online enquêtes in het algemeen te benadrukken, wordt echter (weliswaar in zekere mate terecht) ook gesteld dat de betrouwbaarheid van telefonische enquêtes, en de mate van willekeurigheid van bijbehorende steekproeven, het laatste decennium sterk aan het afnemen is. Dan is het toch op zijn minst merkwaardig dat de online verkregen gegevens juist gewogen worden naar bestanden die via diezelfde telefoon tot stand gekomen zijn. Als telefonische enquêtegegevens immers niet op een voldoende overtuigende wijze gebaseerd zijn op een willekeurig getrokken steekproef, kan het zo zijn dat een “biased” online bestand gecorrigeerd wordt met behulp van een ander “biased” telefonisch bestand.

Ook de Gouw & Groenland (2000) van Blauw Research te Rotterdam, concluderen dat, wanneer een weegmethode toegepast dient te worden, *propensity weighting* in het algemeen de voorkeur verdient boven poststratificatie, met name wanneer naar een referentiebestand gewogen moet worden. Bij hun positieve bevindingen over online enquêteresultaten gewogen door middel van de propensity score methode, verzuimen echter ook zij te verklaren waarom een telefonische steekproef, die zij eerder in het artikel bij wijze van spreken beschouwen als een niet-willekeurig getrokken steekproef (*convenience sample*),

dan toch als goed referentiebestand kan dienen voor de te wegen online steekproefresultaten. Interessant is overigens wel dat bij de vergelijking tussen poststratificatie en propensity weighting die zij uitgevoerd hebben, poststratificatie leidde tot een meer in plaats van minder vertekend bestand. Gezien de summiere informatie in het artikel over de uitgevoerde procedure, kunnen hier echter geen conclusies aan verbonden worden.

Een meer grondige studie naar de bruikbaarheid van de propensity scoring benadering bij het vergelijken van online en telefoon enquêtes is uitgevoerd door Miller en Panjikaran (2001). Zij hebben *propensity weighting* benadering gebruikt om te zien of waargenomen verschillen tussen online en telefonische onderzoek respons verklaard kan worden door de demografische verschillen tussen de respondentgroepen. Daartoe werden gegevens van vier paar (online & telefonische steekproef) landelijke Amerikaanse consumentenonderzoeken gebruikt, die onder meer betrekking hadden op bioscoopbezoek, sportschoenen, persoonlijke mondhygiëne en consumptie van sterke drank. Voor vrijwel alle antwoorden in de vier categorieën, bleek het wegen van gegevens op basis van de *propensity weighting* benadering ineffectief in het verwijderen van de verschillen tussen online en telefonisch verkregen gegevens. In enkele gevallen bleek het uitvoeren van de *propensity weighting* procedure zelfs schadelijk voor de resultaten, zoals in het geval van vragen over mondhygiëne waarbij er pas een significant verschil tussen de twee groepen ontstond ná het wegen.

Voorts stellen Miller en Panjikaran terecht dat het succes van welke manier van wegen dan ook, voor het grootste deel afhankelijk is van de informatie die beschikbaar is. Als het verschil tussen self-selecting online versus telefonische respondent groepen slechts een kwestie is van demografische kenmerken, zou een eenvoudige stratificatie weegmethode voldoende kunnen zijn om twee groepen te vergelijken. Het zijn juist onmeetbare variabelen zoals houding, ervaringen en persoonlijke voorkeuren, die het verschil in resultaten tussen respondent groepen kunnen verklaren.

Een extra punt van aandacht is de vooronderstelling die geldt bij *propensity weighting* volgens de originele theorie, namelijk dat de waargenomen hulpvariabelen (op basis waarvan de propensity score berekend wordt), niet beïnvloed worden door de methode van enquêteren. Dit is een logische vooronderstelling, zeker gezien de originele medische toepassing van deze theorie. De observatie van een medicus of een patiënt man of vrouw is (geslacht als co-variabele) wordt niet beïnvloed door het feit of een patiënt deel uitmaakt van de echte of placebo-groep van onderzoek.

Maar in enquêteonderzoek kunnen de demografische hulpvariabelen echter wel beïnvloed worden door de methode van enquêteren. Zo kunnen respondenten verschillende antwoorden geven op online enquêtes over inkomen en huwelijks staat dan ze bij telefonische enquêtes zouden doen. Bij het corrigeren van online gegevens met behulp van propensity scoring benadering, is het van groot belang te voorkomen dat de weging afhangt van hulpvariabelen die kunnen verschillen per gehanteerde methode van enquêteren.

Onderzoekers van de Zweedse Linkopings Universiteit hebben ook een casestudy uitgevoerd waarbij *propensity weighting* vergeleken werd met andere weegmethoden die toegepast worden bij dataverzameling via het internet (Varedian e.a., 2003). De gegevens van een consumentenonderzoek in vier Scandinavische landen waren zowel met behulp van online als telefonisch enquêtes verkregen. Bij de online enquêtes werd gebruik gemaakt van een vrijwillig internetpanel. Diverse vragen over levensstijl werden in beide enquêtes opgenomen om weging om andere dan demografische variabelen mogelijk te maken. Zowel een model met als hulpvariabelen alleen levensstijlvariabelen, als inclusief demografische variabelen werd geanalyseerd.

De resultaten laten zien dat *propensity weighting* niet leidt tot betere resultaten ten opzichte van het gebruiken van de standaard (stratificatie) weegmethoden. Dit kan volgens de onderzoekers ten eerste te maken hebben met het feit dat reeds een meerderheid van de Scandinavische bevolking internet toegang heeft, waarmee het nut van het vergelijken van online en offline populaties sowieso beperkt is. Ten tweede is de Scandinavische bevolking, binnen elk land, veel meer homogeen dan bijvoorbeeld de bevolking van de Verenigde Staten. Door deze homogeniteit zijn er weinig onderscheidende kenmerken waar te nemen, die als verklaring kunnen dienen voor het al dan niet hebben van toegang tot internet, zo stellen Varedian en Forsman (2003). Zij lichten deze veronderstelling over een meer homogene bevolking in Scandinavische landen echter niet toe met cijfers.

Door het ontbreken van voldoende verklarende variabelen is het moeilijk om de gegevens met behulp van *propensity weighting* benadering te wegen. De auteurs concluderen tenslotte dat de propensity score weighting methode, ondanks de solide mathematische fundamenten waarop ze gebaseerd is, niet in staat is geweest voor een mogelijke bias in online enquêtes te corrigeren.

Een korte analyse van Schonlau e.a. (2001) heeft weliswaar niet direct betrekking op *propensity weighting* als methodiek, maar deelt wel een zelfde uitgangspunt wat betreft de online-offline vergelijking van enquêteresultaten. Het uitgangspunt is namelijk het gebruiken van een (telefonisch) willekeurig-getrokken steekproef om voor de bias in een (online) niet-willekeurig getrokken steekproef te corrigeren. In de analyse wordt door middel van een wiskundige berekening getracht aan te tonen hoe nuttig het is om voor een dergelijke gecombineerde online en offline steekproef te kiezen. Het grote voordeel van een online *convenience sample*, namelijk de lage marginale kosten per respondent, zou zo gecombineerd kunnen worden met de betrouwbaarheid van willekeurig getrokken steekproeven. De auteurs gebruiken als voorbeeld de keus tussen een willekeurig getrokken steekproef van 5.000 individuen, of één van 4.000 individuen maar aangevuld met een *convenience* steekproef van 10.000 individuen tegen dezelfde totale kosten.

Het idee is dat de willekeurig getrokken steekproef gebruikt kan worden om de bias in de *convenience sample* te meten voor elke variabele. Met een schatting van de hoeveelheid bias zou men de informatie van beide steekproeven kunnen combineren om tot een meer precieze schatting te komen dan mogelijk zou zijn op basis van de willekeurig getrokken steekproef alleen. Tevens zou het mogelijk zijn om uitspraken te doen over kleinere sub-groepen door het grotere aantal respondenten.

Na de analyse blijkt echter "*that there is no point in using a probability sample to remove the bias from a convenience sample.*" (Schonlau e.a, 2000 p. 111). Ook het gebruik van een ongecorrigeerde *convenience sample* om als aanvulling te dienen op een willekeurig getrokken steekproef, is onderhevig aan stringente voorwaarden. Vanuit een praktisch oogpunt lijken mogelijkheden tot het samenvoegen van een offline willekeurig getrokken steekproef en een online niet-willekeurig getrokken steekproef dus zeer beperkt, zo betogen de schrijvers.

Toekomst van propensity weighting

Hoewel er wel positieve resultaten geboekt zijn met *propensity weighting* bij online onderzoek, is nog niet duidelijk hoe deze zich verhouden tot resultaten die met traditionele wegingmethoden behaald zouden zijn, ceterus paribus.

Bij het op waarde schatten van *propensity weighting* voor online onderzoek, zijn twee vragen van belang. Ten eerste de vraag of het herwegen van onderzoeksgegevens, verkregen door middel van online onderzoek, met behulp van een telefonische referentiebestand, een

positieve invloed heeft op de betrouwbaarheid van resultaten. Als tweede volgt dan de vraag, of *propensity weighting* de juiste methode is die deze weging kan uitvoeren, en de theoretische uitgangspunten van de methode ook leiden tot meetbaar betere resultaten na herweging van online onderzoeksgegevens.

Zoals in paragraaf 2.3.2 bleek, staan de representativiteit van telefonische steekproeven juist in toenemende mate onder druk. Voor alleen demografische afwijkingen in de groep van telefonische respondenten kan nog wel gecorrigeerd worden door middel van stratificatie, maar met een afnemende response rate kunnen ook in telefonische steekproeven biasen ontstaan. Als immers de groep telefonische non-respondenten groter wordt, kunnen soortgelijke problemen ontstaan die optreden bij online niet-willekeurig getrokken steekproeven, daar veroorzaakt door het *self-selecting* karakter. Telefonische respondenten zijn dan als het ware *self-selecting* telefonische respondenten, omdat ze in tegenstelling tot een grote groep non-respondenten, wél willen meewerken aan de enquête. Het is dus twijfelachtig of telefonische parallel gehouden enquêtes als betrouwbare referentiebron kunnen dienen bij het wegen van online enquêtegegevens. Als ook nog bedacht wordt dat het slechts een kwestie van jaren is voordat internettoegang even wijdverspreid is als telefoonaansluitingen dat nu zijn, lijkt het nut van het wegen van online enquêtegegevens op een telefonisch referentiebestand in de toekomst alleen maar meer aan belang in te moeten winnen.

De tweede vraag is, ervan uitgaande dat het herwegen van online gegevens met een telefonisch referentiebestand nuttig is, of *propensity weighting* als methode daadwerkelijk tot betere resultaten kan leiden. Een belangrijk uitgangspunt voor het herwegen van online gegevens is het corrigeren voor de *self-selection* bias die optreedt bij niet-willekeurig getrokken online steekproeven. Alleen door de verklarende variabelen, die in verband staan met deze *self-selection* bias, te kunnen selecteren en meten, kan een weging uitgevoerd worden die voor de *self-selection* bias corrigeert. Deze verklarende variabelen met betrekking tot het *self-selection* proces van online respondenten zijn echter nog nauwelijks in kaart gebracht, en het is ook de vraag in hoeverre dit gaat gebeuren gezien het complexe karakter van dergelijke individuele beslissingsprocessen.

Op basis van de, weliswaar nog prille, ervaringen van wetenschappelijke onderzoeken waarover gepubliceerd is, kunnen dus bij beide vragen vraagtekens gezet worden. Verder onderzoek is nodig voordat bekend wordt of *propensity weighting*, als methodiek voor het wegen van online enquêteresultaten, haar experimentele karakter achter zich zal kunnen laten.

3.3 Ontwikkeling van Internet Panels

Een onderzoekspanel bestaat uit respondenten die aangegeven hebben mee te willen werken aan een of meerdere toekomstige onderzoeken. Doorgaans zijn vooraf al de meest belangrijke demografisch variabelen bekend, waarvan gebruik kan worden gemaakt bij de steekproef-trekking of onderzoeksopzet.

Het fenomeen van panels is niet nieuw, reeds vanaf de jaren 80 jaren hebben vooral consumenten-panels een significante bijdrage geleverd aan het verzamelen van kennis op het gebied van marketing (Dreze e.a., 1998). Er zijn twee belangrijke voordelen te noemen die het gebruik van panels aantrekkelijk maakt, ten eerste is dat de eerder genoemde beschikbaarheid van de vooraf verzamelde kenmerken van respondenten. Ten tweede bestaat de mogelijkheid om respondenten over een langere periode te volgen, waarmee longitudinale studies mogelijk worden. Door het volgen van individuele respondenten kan de invloed van bijvoorbeeld promotie op het aankoopgedrag van consumenten beter ingeschat worden.

De traditionele panels kwamen echter altijd tot stand door willekeurig getrokken steekproeven. Het is juist hierin dat internetpanels verschillen van opzet, want internet panelrespondenten zijn vaak via *self-recruiting* methoden geworven. Wat de gevolgen daarvan voor de kwaliteit van een dergelijk internetpanel zijn, zal in navolgende paragraaf besproken worden.

3.3.1 Internetpanels

Het uitvoeren van online onderzoek met behulp van een internetpanel is een combinatie die de laatste jaren veel aan populariteit heeft gewonnen. Eerst waren het de kleine, nieuwe onderzoeksbureaus die experimenteerden met internetpanels, terwijl pas later de traditionele onderzoeksbureaus volgden die aanvankelijk moeilijk de omslag naar online onderzoek wisten te maken (Emerce, 2001).

Het opbouwen en onderhouden van een panel brengt een aanzienlijke investering met zich mee (Grossnickle & Raskin, 1999). Voor het kunnen bedienen van klanten uit veel verschillende disciplines is een panelgrootte van tenminste enkele tienduizenden respondenten nodig. Inmiddels hebben zich zowel in Nederland als Amerika een beperkt aantal grote onderzoekspanels gevormd, die meestal direct danwel indirect verbonden zijn aan grotere onderzoeksbureaus. Binnen Europa hebben diverse samenwerkingsverbanden tevens geleid tot gestandaardiseerde Europese panels.

Er zijn grofweg twee redenen waarom internetpanels zo populair zijn bij het doen van online onderzoek. De eerste ligt verscholen in het feit dat een makkelijke en betrouwbare willekeurig steekproeftrekking met behulp van RDD niet mogelijk is op het internet. Omdat met het werven, benaderen en selecteren van respondenten via internet relatief veel meer kosten, tijd en moeite gemoeid zijn indien de respondentengroep geacht wordt in zekere mate representatief te zijn voor een grotere populatie, wordt het vanuit bedrijfskundig oogpunt aantrekkelijk om dezelfde respondentengroep te gebruiken voor meerdere onderzoeken. Een internetpanel biedt dan uitkomst. Ten tweede blijkt panelonderzoek zeer effectief bij het doen van doelgroeponderzoek. Door te beschikken over een grote database met gegevens van respondenten, kunnen specifieke doelgroepen (bijvoorbeeld chronisch zieken, tieners, etcetera) makkelijk geselecteerd worden voor onderzoek. Tot slot hebben voordelen van online onderzoek in het algemeen (zie hoofdstuk 2.4) ook bijgedragen aan de populariteit van internetpanels.

Werving van respondenten

Het bouwen en onderhouden van een internetpanel is mogelijk door respondenten te werven door middel van een willekeurig getrokken steekproef of een niet-willekeurig getrokken steekproef, of een combinatie hiervan. Bij de laatste twee, niet-willekeurig getrokken steekproefrespondenten, gaat het meestal om zogenaamde *self-recruited* respondenten. Zij hebben zichzelf als het ware geselecteerd voor het panel, door te reageren op een algemene oproep of advertentie.

Vrijwel alle bekende internetpanels zijn gebaseerd op een combinatie van wervingsmethoden, waarbij *self-recruited* respondenten in aantallen het belangrijkste zijn. In paragraaf 2.5.1 bleek al dat deze internetpanels meer dan andere typen niet-willekeurig online getrokken steekproeven in staat zijn om een representatieve steekproef te benaderen (Alvarez 2000, Grossnickle & Raskin 1999).

Omdat een panel als basis dient voor vele verschillende onderzoeken, is de manier waarop een panel opgebouwd wordt van cruciaal belang voor de betrouwbaarheid van het panel, des te meer als het panel ook als basis dient voor onderzoeksresultaten die gegeneraliseerd worden naar een grotere populatie.

Een rondgang langs de websites van onderzoeksbureaus die een groot panel beheren, leert dat juist op dit punt, de werving van respondenten, de bureaus grote voorzichtigheid betrachten wat betreft het verstrekken van informatie. Het blijft meestal bij uitspraken als:

“Our proprietary recruiting techniques include a number of different online and offline methods.” (bron: www.metrixlab.nl)

Eenzijds houdt deze terughoudendheid verband met het feit dat een internetpanel behoort tot de kernbezittingen van een onderzoeksbureau. Door een te groot inzicht vrij te geven in de methodiek waarmee een internetpanel tot stand is gekomen, kunnen ook zwakke punten van het internetpanel aan het licht gebracht worden.

Er bestaat immers een tegenstelling tussen de betrouwbaarheid van de methode of het kanaal waarop respondenten geworven worden, en de kosten en moeite die aan die werving verbonden zijn. Zo is het bijvoorbeeld makkelijk om aan lijsten met e-mail adressen te komen, maar in hoeverre deze betrouwbaar of representatief zijn is meestal onbekend. Adverteren op websites die maar een kleine doelgroep bedienen hoeft niet veel te kosten, maar juist de websites die een zeer groot (en dus meer representatief) bereik hebben zoals zogenaamde *portals* en zoekmachines hebben hogere *click-through* tarieven. Door het aanbieden van prijzen of beloningen is het makkelijker om respondenten te werven, maar het is waarschijnlijk dat er dan tevens een grotere bias in een panel geïntroduceerd wordt.

Een dergelijke bias is gevolg van het proces van *self-selection* van respondenten in de werving. Zij treedt in meer of mindere mate op bij vrijwel alle directe werving via internet van respondenten. Door gebruik te maken van een parallelle additionele telefonische steekproef kan getracht worden voor deze bias te corrigeren, zoals we in het vorige hoofdstuk hebben kunnen zien. Mocht dit niet tot de mogelijkheden behoren, dan rest doorgaans alleen de methodologie van pre- en poststratificatie om in ieder geval voor de sociaal-demografische component van de bias te kunnen corrigeren en een zekere mate van representativiteit te kunnen claimen.

Deze werving van respondenten op het internet kan op meerdere manieren plaatsvinden. Het meest gebruikelijk is het plaatsen van advertenties op websites, met behulp van zogenaamde *banners* of *pop-up windows*. In elk geval is het belangrijk op een verzameling van websites te adverteren waarvan zo'n groot mogelijk gedeelte van de internet populatie bezoekt.

De traditionele *banner* is een reclame-blok die op een webpagina geplaatst wordt, waarmee de respondent door erop te klikken door kan surfen naar de pagina van de adverteerder. De *banner* heeft de laatste jaren flink aan effectiviteit ingeboet, was er in 1997 bijvoorbeeld nog een *click-through rate* van boven de 5% mogelijk, tegenwoordig neemt nog geen eens 0,5% van de bezoekers de moeite om door te klikken (Multiscope, 2003). Toch blijft de *banner* de meest gebruikelijke manier van adverteren op het internet, alleen door de zeer lage respons lijkt ze steeds minder geschikt voor de werving van panel-respondenten. Een paar jaar geleden leek de *pop-up window*, een venster dat automatisch vóór of achter de browser verschijnt bij het bekijken van een webpagina, opnieuw in staat om de aandacht van de internetgebruiker te trekken. Ook daar maakt interesse echter al snel plaats voor irritatie bij de internetgebruikers, en nu Microsoft standaard een *pop-up blocker* implementeert in haar browser, lijkt ook voor de *pop-up* geen gouden toekomst weggelegd. De laatste trend in internet reclame zijn advertenties die tijdelijk controle over het scherm lijken te nemen, door over de te lezen webpagina heen (“*overlay*”), geanimeerde afbeeldingen te tonen. Ook hier is het waarschijnlijk een kwestie van tijd voordat intelligente *blockers* ontwikkeld zijn.

De adverteerder moet steeds nieuwere internettechnologieën op een creatieve manier gebruiken om de consument te verrassen. Maar zelfs als dat gebeurt, is het de vraag of de dalende trend in responspercentages van online advertenties tot stoppen kan worden gebracht (Multiscope, 2003). Met deze responspercentages valt of staat echter niet de werving van respondenten, die plaats kan blijven vinden hetzij door (bij een lagere effectiviteit) meer advertenties uit te zetten en een mogelijk grotere selectiviteit van non-respons voor lief te nemen, hetzij door het vinden van nieuwe kanalen waarmee de internetgebruiker bereikt kan worden.

Onder andere mogelijkheden om panel-respondenten te werven valt bijvoorbeeld het aangaan van samenwerking met andere internetbedrijven, die om uiteenlopende redenen een registratiesysteem kennen. Door *co-registration* kunnen gegevens van registrerende internetgebruikers doorgespeeld worden aan bijvoorbeeld een “*partner-network*” onderzoeksbureau. Er valt dan te denken aan bijvoorbeeld registratie voor (gratis) internetgebruik, registratie bij online media (kranten) of registratie voor andere online diensten. Men kan ook bestaande leden van een internetpanel aanmoedigen om nieuwe leden te werven. Wat echter de echte succesvolle communicatie en marketing kanalen van de toekomst worden met betrekking tot de werving van panel-respondenten, valt nog te bezien.

Wat betreft het onderhouden van een bestaand panel zijn twee opmerkingen te plaatsen. Allereerst moet er rekening gehouden worden met een zogenaamde panelmoeheid die kan optreden als respondenten zich gaan gedragen als ervaren enquête-invullers. Dit kan voorkomen worden door het aantal enquêtes waaraan een respondent kan meewerken te beperken, terwijl het middels een eenvoudige analyse mogelijk is om de ergste uitschieters te herkennen. Een analyse naar panelmoeheid bij het Knowledge Networks panel (Dennis, 2003) liet zien dat enquêteresultaten maar zeer beperkt beïnvloed werden door verschillen tussen antwoorden van nieuwe en ervaren respondenten.

Ten tweede zien we dat door informatie of zogenaamde *infotainment* aan te bieden op een website het mogelijk is de respondenten meer betrokken te laten zijn bij het proces van enquêteren en de opdrachtgevers. Een aardig voorbeeld hiervan is de website www.opinieland.nl, de portaal voor het internetpanel van Bloomerco. Deze website biedt behalve achtergrondinformatie over uit te voeren onderzoeken ook mogelijkheden tot communiceren met andere leden van het panel. Deze vorm van *community building* kan een positief effect hebben op de betrokkenheid van de panelleden, alhoewel men zich kan afvragen of een dergelijke *community* niet kan bijdragen aan een nieuwe, onbekende bias component.

Panel op basis willekeurig getrokken steekproef

Tot nu toe bestaat er, zover bekend, maar één bedrijf wat op grote schaal een internetpanel beheert wat geheel gebaseerd is op willekeurig getrokken steekproeven. Het Amerikaanse bedrijf *Knowledge Networks* beheert een internetpanel dat geheel tot stand gekomen is via RDD-steekproefmethoden. Het panel bevat zowel gebruikers als niet-gebruikers van internet. Dit is mogelijk doordat het bedrijf de apparatuur installeert die aangesloten kan worden op de televisie en toegang biedt tot internet via de telefoonlijn. Doordat het onderzoeksbureau de kosten van zowel de apparatuur als de aansluiting voor haar rekening neemt, is het mogelijk om een bredere doorsnede van de bevolking te bereiken dan anders mogelijk zou zijn.

Knowledge Networks claimt in een publicatie (Dennis, 2003) dat haar panel superieur is boven andere internetpanels, en ondersteunt dat met een vergelijkende studie van Krosnick & Chang, waarbij eenzelfde enquête voorgelegd wordt aan een willekeurig getrokken telefonische steekproef, het *Knowledge Networks* panel en een internet panel van concurrent *Harris Interactive*, die met name op basis van vrijwillige internetwerving tot stand gekomen is. Het is het internetpanel van *Knowledge Networks* dat positief uit de vergelijking komt. Als kanttekening bij het onderzoek kan wel gesteld worden dat de vergelijking heeft plaatsgevonden op basis van ruwe, ongecorrigeerde cijfers van de concurrent. *Harris Interactive* hecht echter veel waarde aan haar “*propensity weighting*” techniek, die voor de waargenomen bias in antwoorden zou moeten corrigeren. Toch neemt dat niet weg dat enquêtegegevens die van nature, vóór weging, meer representatief zijn altijd een betere uitgangspositie bieden.

Een onderzoek van Smith (2001) vergelijkt een enquête uitgevoerd door *Knowledge Networks* met de enquêteresultaten van een grootschalig onderzoek van het *National Opinion Research Center* in het kader van de *General Social Survey 2000*. Dit laatste onderzoek is uitgevoerd met behulp van persoonlijk afgenomen enquêtes bij ruim 2000 Amerikaanse huishoudens. Allereerst blijkt uit de vergelijking dat bij beide onderzoeken de resultaten vóór weging niet veel verschillen van de resultaten ná weging. Dit houdt mogelijk verband met een goede representativiteit van de steekproef, waardoor er weinig ondervertegenwoordiging optreedt waarvoor gecorrigeerd (gewogen) zou moeten worden. Verder lijken de resultaten van de twee onderzoeken op de meeste variabelen goed te vergelijken. De weinige verschillen die wel opvallen, zijn vooral te wijten aan het verschil in de enquêtemethodiek, namelijk persoonlijk (mondeling) versus internet (visueel). Zo wordt bijvoorbeeld in het persoonlijk afgenomen onderzoek veel meer steun toegezegd aan arme achterstandswijken dan in het online onderzoek. De persoonlijke manier van enquêteren leidt mogelijk tot sociaal wenselijke antwoorden. Het onderzoek van Smith (2001) laat zien dat met twee onderzoeken die beide gebaseerd zijn op een willekeurig getrokken steekproef, het niet zonder meer duidelijk is welke vorm van enquêteren tot betere resultaten leidt.

In een onderzoek van Berrens (in Schonlau e.a., 2001), wordt eenzelfde enquête voorgelegd aan respondenten uit zowel het *Knowledge Networks* en *Harris Interactive* internetpanel als uit een (RDD) telefonische steekproef. Het blijkt dat na weging op demografische variabelen, alle drie de onderzoeken statistisch gelijke resultaten opleveren op de meeste onderzoeksvragen.

Meer onafhankelijk onderzoek is nodig, voordat er met grotere zekerheid uitspraken kunnen worden gedaan over hoe internetpanels die geselecteerd zijn op basis van een willekeurige RDD-steekproef, zich verhouden tot zowel traditionele telefonische enquêtes en internetpanels waarbij door middel van weging gecorrigeerd is voor het niet-willekeurige karakter van de werving. Dit kan ook verschillen al naar gelang het uitgangspunt en de opzet van een onderzoek, waarover nu meer.

3.3.2 Bruikbaarheid van online panels

Internet panelonderzoek heeft inmiddels een stevige plaats verworven op het gebied van online onderzoek. In deze conclusie zal het nut van internetpanels bekeken worden vanuit twee verschillende invalshoeken, die betrekking hebben op de aard van een onderzoek. Door dit onderscheid te maken kan een meer genuanceerd beeld worden geschapen met betrekking tot de bruikbaarheid van online panels.

Focus op trends & ontwikkelingen in populatie of op specifieke doelgroepen

Als er kritiek komt op marktonderzoeken over het wetenschappelijke gehalte van een bepaalde publicatie, wordt hierop vaak gereageerd met een praktisch ingesteld antwoord. De managers of beleidsbeslissers, die opdracht gegeven hebben tot de onderzoeksrapportage, maken naar aanleiding van de publicatie toch dezelfde zakelijke beslissingen. Of er bij wijze van spreken een stijging van 4 of 6 procent is maakt niet zoveel uit, de beslissing blijft hetzelfde (Multiscope, 2003). En een groot deel van het marktonderzoek vindt inderdaad plaats omdat bepaalde ontwikkelingen of trends in kaart gebracht moeten worden. Er moet een inzicht gegeven worden in ontwikkelingen en mogelijke verklarende factoren, maar dat behoeft niet perse te gebeuren aan de hand van nauwkeurig geschatte, statistisch significante parameters. De verhouding tussen verschillende geschatte parameters geeft vaak nuttiger informatie voor de opdrachtgever dan de absolute waarden daarvan. Als het doel van een onderzoek is om de rangorde te kennen van een aantal factoren of waardering van producten, kan verwacht worden dat internet panelonderzoek en traditioneel onderzoek tot gelijke zakelijke beslissingen zullen leiden (Miller, 2001).

Dat juist bij het in kaart brengen van trends en ontwikkelingen internetpanels een nuttige rol kunnen vervullen, heeft te maken met het feit dat problemen met de self-selecting aard van dataverzameling hierbij een minder grote invloed hebben op de resultaten. Hoewel de mogelijke self-selecting bias die bestaat bij online enquêteresultaten zeker niet als constant mag worden verondersteld, zal in ieder geval de richting van die bias binnen een kort tijdsbestek niet snel veranderen. Dus waar de absolute waarden van variabelen wel sterk beïnvloed worden door de bias, gebeurt dat in ieder geval een stuk minder met de relatieve waarden tussen de variabelen, oftewel trends en ontwikkelingen. Ervan uitgaande dat we te maken hebben met een internetpanel waarbij zorgvuldig aandacht is besteed aan de samenstelling en eventuele weging van enquêteresultaten, kan gesteld worden dat, waar het gaat om het signaleren van ontwikkelingen of trends, een internetpanel betrouwbare onderzoeksresultaten kan leveren.

Een tweede onderzoeksgebied waar de mogelijkheden van een groot internetpanel goed tot zijn recht komen, is daar waar specifieke doelgroepen het onderwerp van onderzoek zijn. Omdat bij internetpanels standaard al veel sociaal-demografische informatie bekend is over de respondenten, is het vrij eenvoudig een steekproefkader vast te stellen waarin alleen respondenten voorkomen die van belang zijn bij het onderzoek. En omdat de grotere internetpanels doorgaans minstens enkele tienduizenden respondenten bevatten, is het ook bij specialistische doelgroepen mogelijk om een voldoende steekproefgrootte te realiseren. Doelgroeponderzoek is een type onderzoek waar bij uitstek een internetpanel geschikt voor is, doordat het snel en tegen lage kosten voor een efficiënte dataverzameling kan zorgen. Wel blijft het belangrijk rekening te houden met mogelijke effecten van de *self-selection* bias op absolute generalisaties van resultaten naar een grotere populatie, een belangrijke kwestie waarmee we dit hoofdstuk besluiten.

Focus op absolute generalisaties naar populatie

Uiteindelijk komen alle resultaten, ook zij die betrekking hebben op trends en ontwikkelingen zoals die hiervoor besproken zijn, voort uit gegevens die verkregen zijn uit de antwoorden van ingevulde enquêtes. Absolute kengetallen, die gegeneraliseerd kunnen worden naar een grotere populatie, staan het dichtst bij de brongegevens en vormen vaak ook de meest interessante en makkelijk te interpreteren resultaten. Niet voor niets wordt bij de bespreking van een onderzoek in de krant vrijwel altijd gerefereerd aan de absolute percentages⁵ ("57 procent van de Nederlanders brengt iedere week een bezoek aan de....").

De vraag of op basis van internetpanels betrouwbare generalisaties naar een grotere populatie gemaakt kunnen worden, is een van de kernvragen bij het debat over de betrouwbaarheid van online onderzoek. Vooralsnog overheersen hierbij de vraagtekens, wellicht mede omdat het fenomeen van internetpanels, en de bijbehorende methodologie, nog zo jong is.

Een Zweedse studie maakte bij het beantwoorden van de vraag onderscheid tussen generalisatie van resultaten naar de *internetpopulatie*, en naar de *totale populatie* (dus inclusief niet-internetgebruikers). De studie, uitgevoerd door Nyfjall (2001), baseerde zich op twee parallelle enquêtes, waarbij het doel was te onderzoeken of een *self-selected* internetpanel gebruikt kan worden om conclusies te trekken over een grote algemene populatie of over de populatie van internetgebruikers. De conclusie gaf aan dat internetpanels redelijk goed gebruikt kunnen worden om tenminste een indicatie te geven van de werkelijke waarden die voorkomen in een internetpopulatie. Mogelijkheden tot betrouwbare generalisaties over het algemene publiek, gebaseerd op een internetpanel, zijn echter veel meer beperkt, zo bleek uit zijn onderzoek. Zeker als het om vragen gaat die betrekking hebben op de technologische innovaties, interesse of de penetratie van informatie- en communicatietechnologie (ICT) toepassingen blijken internetpanels ongeschikt om generalisaties naar het algemene publiek te maken. Het onderscheid wat in de studie van Nyfjall gemaakt wordt, namelijk tussen de internetpopulatie en de totale populatie, wordt echter minder belangrijk naarmate de dekking van internet toeneemt.

Een potentiële spelbreker waarvoor minder makkelijk een oplossing te bedenken is, heeft betrekking op de combinatie van een niet-willekeurig getrokken steekproef en het proces van *self-selection* bij de werving van internetpanel-respondenten. Meer vergelijkende studies tussen willekeurig getrokken en niet-willekeurig getrokken steekproeven moeten plaatsvinden voordat vragen omtrent de generaliseerbaarheid van *self-selecting* internetpanels beter beantwoord kunnen worden (Sheehan, 2002).

Met een telefonische steekproef die nu sowieso nog het meest representatief is voor de algemene populatie lijkt een op RDD-gebaseerd internetpanel een veel betere uitgangssituatie te bieden dan een internetpanel met *self-selecting* respondenten. Hoe groot de impact kan zijn van het feit dat respondenten zichzelf selecteren in een vrijwillig internetpanel, kan mooi worden geïllustreerd met onderstaande quote van marktonderzoek journalisten Grossnickle & Raskin (1999, p. 2):

"RDD is good, approaches that rely on panel members proactively seeking out and joining panels are generally more suspect, because you can end up with a bunch of people who have nothing better to do with their time than take survey after survey in the hope of one day winning a fabulous cruise to Cabo San Lucas"

⁵ Percentages zijn in principe altijd *relatief*, met *absoluut* wordt in deze context bedoeld dat er een meting van de verdeling van één variabele op één punt in de tijd wordt weergegeven, zonder dat die gerelateerd wordt aan andere variabelen of tijdsperioden.

3.4 *Blik op de toekomst*

We hebben kunnen zien dat het nadeel wat online onderzoek heeft met betrekking tot de gebrekkige dekking ten opzicht van telefonische en per post verstuurd enquêtes in sneltempo aan het verdwijnen is door de snelle penetratie van internet. De kern van de problematiek waar steeds op teruggekomen wordt, zit hem in het feit dat een random getrokken steekproef aan de basis staat van nauwkeurig steekproefonderzoek in het algemeen, en in het bijzonder daar waar generalisaties naar een grotere populatie gemaakt moeten kunnen worden. Was dat random trekken van een steekproef via de post en telefoon nog redelijk goed mogelijk (RDD), bij internet zijn hiervoor geen indexen met contactgegevens van internetgebruikers beschikbaar. Of die op de middellange of lange termijn wellicht wel beschikbaar kunnen gaan worden is vervolgens de vraag. Waarschijnlijk vanwege het hoge "koffiedikkijk"-gehalte zijn hierover geen voorspellingen terug te vinden in de literatuur over online onderzoek.

Technisch is het echter wel degelijk mogelijk de totale populatie van internetgebruikers op te nemen in een index, van waaruit een random steekproef getrokken kan worden. Als ook contactgegevens beschikbaar zijn kunnen respondenten vervolgens aangeschreven worden. Concreet zou dat er nu op neer komen dat alle internet service providers, die toegang verstrekken aan thuisgebruikers van internet, een wettelijk bepaalde hoeveelheid informatie verstrekken aan een (overheids)-instituut die deze index zou gaan beheren.

Een tweede mogelijkheid, wellicht op de iets langere termijn, is gebruik te maken van een uniek adres wat alle computers aangesloten op het internet reeds hebben, namelijk het zogenaamde ip-adres. Net als telefoonnummers zijn ook ip-nummers volgens een bepaalde systematiek opgebouwd. Zo beginnen alle ip-nummers van computers aangesloten op het universiteits netwerk van Utrecht met *131.211.x.x*, waarbij x staat voor nog tweemaal een getal tussen de nul en de duizend (voorbeeld ip-adres: *131.211.70.21*). Het zou theoretisch in de toekomst mogelijk zijn om, net als de RDD-techniek, random ip-adressen te selecteren binnen een bepaalde *range* (bijvoorbeeld: Nederland) en vervolgens via deze selectie aan ip-adressen op de bijbehorende computers een uitnodiging weer te geven tot deelname aan onderzoek.

De techniek an sich is dus niet het grootste probleem, het is meer de vraag of de samenleving het bestaan van dergelijke mogelijkheden zal willen accepteren. Het toenemende gebruik van computers en netwerken in de samenleving is hand in hand gegaan met een toenemende bezorgdheid van mensen over het behoud van hun privacy. Mensen verlangen een grotere controle over het al dan niet openbaar maken van hun eigen gegevens, en accepteren steeds minder makkelijk dat onbekenden (veelal vanuit *direct marketing* organisaties) ongevraagd contact opnemen met hen.

Dat deze ontwikkelingen ook in wetgeving verankerd worden, laat het eerder besproken initiatief van de overheid van de Verenigde Staten zien, waarbij alle inwoners van de Verenigde Staten hun telefoonnummer kunnen noteren op de website www.donotcall.gov, waarna *direct marketing* bedrijven deze nummers niet meer mogen bellen op straffe van aanzienlijke boetes. Maar terwijl de consumenten meer mogelijkheden krijgen om adverteerders te weren, zitten de adverteerders op hun beurt echter ook niet stil. Mooi voorbeeld hiervan is de recente trend van zogenaamde "*spy-ware*", software die ongemerkt op computers van consumenten geïnstalleerd wordt bij het installeren van bepaalde gratis verkrijgbare computerprogramma's. Dit kleine stukje software werkt als een soort spion, en bestudeert de computer en de gebruiker bij de dagelijkse activiteiten, en tovert op gezette tijden reclame boodschappen op het scherm. Is bijvoorbeeld de homepage van een automerk bezocht, dan zorgt de software-spion ervoor dat de surfer steeds advertenties van de fabrikant krijgt te zien (Volkskrant, 2003).

Tot nu toe heeft het internet zich geëvolueerd tot een wereldwijd netwerk waar weinig controle of toezicht plaatsvindt. De internetgebruiker voelt zich redelijk anoniem, en het grensoverschrijdende karakter van internet vormt het grootste obstakel tot effectieve controle over publicaties en berichtenverkeer op het internet, nog afgezien van de vraag of daar behoefte aan is. Toch is het nog maar de vraag of de schijnbare anonimiteit van het internet wel zal blijven bestaan. Nu al ondervinden grote sectoren van de economie, zoals de muziek- en filmindustrie, grote problemen van de ongebreidelde kopieerdrijf van "downloaders" die ervoor gezorgd heeft dat er al sinds enige jaren een dalende trend in muziekverkopen plaatsvindt.

Er bestaat dus een strijd tussen verschillende groepen in de samenleving die elke hun eigen visie hebben op de mogelijkheden en beperkingen van internetgebruik. De toekomst zal uit moeten wijzen naar welke kant de balans door zal slaan. Indien het internet in de toekomst meer gereguleerd zal worden door diverse (supra)nationale instituties, zal het wellicht ook mogelijk worden om onder bepaalde voorwaarden een random steekproef van internetgebruikers te kunnen trekken. Voor de nabije toekomst is dit scenario echter niet waarschijnlijk.

4 Casestudy: Drie onderzoeken vergeleken

4.1 Inleiding

In de voorgaande hoofdstukken is gebleken dat online onderzoek mogelijk is in veel verschillende vormen, waarbij het gebruik van een internetpanels het meest populair is onder de grotere onderzoeksbureaus. Maar ondanks een groeiend aandeel van online onderzoek, worden traditionele onderzoeksmethoden nog steeds als superieur beschouwd als het gaat om de betrouwbaarheid en representativiteit.

Om enig inzicht te krijgen in hoeverre resultaten verschillen als gevolg van een andere (online danwel offline) methodiek, zullen in dit hoofdstuk drie verschillende Nederlandse onderzoeken over eenzelfde onderwerp naast elkaar gelegd worden. Daarbij zal de samenhang tussen het uitgangspunt van elk onderzoek en de uiteindelijke resultaten vergeleken worden.

Er is voor gekozen om onderzoeken over "e-commerce" als onderwerp te nemen. Ten eerste omdat er door verschillende grote onderzoeksbureaus over eenzelfde periode onderzoek gepleegd is naar dit onderwerp en het breed in de belangstelling staat. Ten tweede omdat juist dit onderwerp zich goed leent voor online onderzoek, daar het onderwerp zelf met name, maar niet uitsluitend, betrekking heeft op internetgebruikers.

Er wordt niet beoogd een diepgaande analyse uit te voeren over de juistheid van de gegevens van de drie onderzoeken, enerzijds omdat daarvoor gedetailleerde gegevens ontbreken en anderzijds omdat dit ook buiten het doel van deze scriptie valt. Er zal volstaan worden met een korte vergelijking van de opzet en resultaten van deze e-commerce onderzoeken. Het meest interessante en onderscheidende gegeven hierbij is het feit dat twee van deze onderzoeken gebaseerd zijn op telefonische enquêtes, terwijl bij het derde onderzoek door middel van online enquêtes data zijn verzameld.

Er zal een korte beschrijving gegeven worden van de volgende onderzoeken:

- *NIPO's "Wereldwijd E-commerce Onderzoek" (editie 2000 & 2001)*
- *CBS "De Digitale Economie" (editie 2002 & 2003)*
- *Multiscope's "E-commerce Nederland" (editie 2000 & 2001)*

Genoemde jaren zijn de jaren die vermeld worden in de titel (NIPO & CBS) of op het voorblad (Multiscope) van de publicatie. Zij hebben altijd betrekking op het jaar waarop de onderzoeksresultaten van toepassing zijn, het Centraal Bureau voor de Statistiek (CBS) publiceert haar cijfers over internetgebruik in 2000 (2001) pas in de publicatie "De Digitale Economie 2002 (2003)".

Voor deze vergelijking is gebruik gemaakt van onderzoeken die betrekking hebben op het jaar 2000 en 2001. Het was de tijd dat de internethype haar piek bereikte, en er in alle drukte een grote behoefte was aan gegevens en achtergrondinformatie over de internetgebruiker. Al deze drie onderzoeken trachten te voldoen aan deze behoefte aan informatie. Kenmerkend is dat het enige niet-commerciële onderzoeksbureau, het CBS, met een vertraging van bijna 2 jaar haar resultaten publiceerde, terwijl haar commerciële concurrenten gegevens die in het voorjaar verzameld waren reeds in het najaar publiceerden. Dit kwam echter ook deels voort uit de andere systematiek die het CBS hanteert bij de dataverzameling.

Aanvankelijk was het de bedoeling ook een vierde vergelijkbaar onderzoek, de "Pan European Internet Monitor 2001" van onderzoeksbureau Pro Active mee te nemen in de vergelijking. Dit onderzoeksbureau, dat in Nederland als een van de pioniers van online

onderzoek gezien kan worden, is echter inmiddels failliet verklaard (Webwereld, 2003). De desbetreffende onderzoeksrapportages konden daardoor niet meer achterhaald worden.

Voordat een vergelijking van resultaten plaats zal vinden, zal allereerst in de hieropvolgende paragraaf meer inzicht geboden worden in de opzet van elk onderzoek, in zoverre daar informatie over beschikbaar is. De verwachting bestaat dat de onderzoeksopzet en de enquêtemethode in het bijzonder, van invloed zullen zijn op eventueel waar te nemen verschillen in de resultaten. In de derde paragraaf van dit hoofdstuk zullen een drietal uitkomsten van de onderzoeksresultaten met elkaar vergeleken worden. Tot slot zal in de afsluiting van dit hoofdstuk duidelijk worden of er enige conclusies aan deze vergelijking verbonden kunnen worden.

4.2 Onderzoeksopzet

NIPO's Wereldwijd E-commerce Onderzoek.

Dit onderzoek is uitgevoerd in samenwerking met de Taylor Nelson Sofres Group, die een wereldwijd e-commerce onderzoek heeft geïnitieerd en gecoördineerd in zesendertig landen. Het onderzoek vindt vanaf 2000 plaats volgens gestandaardiseerde vragenlijsten zodat onderlinge vergelijkingen tussen landen mogelijk zijn.

NIPO Interactive heeft het Nederlandse deel voor zijn rekening genomen, en daartoe zijn in het voorjaar van 2000 en 2001 telefonische gesprekken afgenomen met personen van 16 jaar en ouder. In 2001 zijn in februari 1574 telefonische gesprekken afgenomen, over 2000 is het precieze aantal respondenten niet bekend.

Voor de hierna volgende vergelijking is het van belang een tweetal definities te kennen, zoals die toegepast zijn op het onderzoek van NIPO.

- *Internetgebruiker*: persoon die in de afgelopen maand, voor één of meer toepassingen, gebruik heeft gemaakt van internet (bijv. surfen, e-mailen, etc.)
- *Online shopper*: internetgebruiker die in de afgelopen maand producten of diensten gekocht of besteld heeft via internet

CBS De Digitale Economie

De CBS stelt sinds 2001 een publicatie “De Digitale Economie” samen, waarin gegevens met betrekking tot de zogenaamde digitale economie afkomstig van een aantal verschillende bronnen samengevoegd zijn. De publicaties zijn het gevolg van een opdracht van het Ministerie van Economische Zaken, waar behoefte bestaat aan gefundeerde gegevens over de relatie tussen ICT en de economie.

De resultaten die voor deze vergelijking bestudeerd worden zijn samengesteld uit uitkomsten van de zogenaamde POLS enquêtes (Permanent Onderzoek LeefSituatie) die continu afgenomen worden onder een grote groep Nederlanders. De vragen over internetgebruik en elektronisch winkelen maken deel uit van de basisenquête, wat wil zeggen dat zo'n 37.000 mensen per jaar deze vragen beantwoorden. Dit gebeurt met behulp van telefonische enquêtes, die in maandelijks evenredige delen gerealiseerd worden. De steekproef wordt uit een personenregister getrokken.

Pre-stratificatie vindt plaats om te voorkomen dat er per huishouden meer dan één persoon deel uit kan maken van de steekproef. Na dataverzameling wordt er gewogen naar het totaal van de bevolking, waardoor de respons tot 15,5 mln. personen wordt opgehoogd. Na deze weging zijn er dus als het ware 15,5 mln. cases, dus de populatieparameters kunnen direct berekend worden (Veen, 2003).

Tot slot de relevante enquêtevragen zoals die door het CBS geformuleerd zijn:

- *Internetgebruiker*: Heeft uw huishouden via de pc een internetaansluiting? (ja/nee)
- *Elektronisch winkelen*: Welke activiteiten worden door uw huishouden gedaan via het internet? (surfen/e-mailen/software downloaden/elektronisch winkelen)

Multiscope E-Commerce Onderzoek

Multiscope, een relatief jong onderzoeksbureau uit 's-Hertogenbosch, is als één van de eerste bureaus begonnen met het doen van onderzoek naar het profiel van de internetgebruiker. Voor het verzamelen van gegevens maakte zij gebruik van online dataverzameling. Het eerste onderzoek naar de internetgebruiker in zijn algemeen vond plaats in 1997, en van 1999 tot en met 2001 voerde Multiscope jaarlijks een onderzoek uit naar e-commerce in Nederland.

De twee relevante onderzoeken uit 2000 en 2001 zijn gehouden onder Nederlandse internetgebruikers van 12 jaar en ouder. Een online enquête is op de Multiscope website geplaatst, en werving heeft plaatsgevonden door het plaatsen van advertenties op websites en door het benaderen van het internetpanel waarover Multiscope beschikt. Voor het valideren van gegevens is in 2000 nog een klein aantal telefonische enquêtes afgenomen, deze zijn echter niet gebruikt in de uiteindelijke analyse en onderzoeksresultaten.

In tabel 4.1 wordt een overzicht van de opbouw van de steekproef gegeven.

Tabel 4.1 Overzicht respons steekproef Multiscope

Jaar	Panel-respondenten			Website werving		Totale steekproef
	benaderd	respons	respons%	soort	respons	
2000	5.489	1.229	22,4%	Banners	440	1.669
2001	3.000	401	13,4%	Pop-ups	1.488	1.889

Bron: Multiscope (2003)

De internetpanel-respondenten zijn eerder via verschillende kanalen geworven, waarbij vrijwel uitsluitend internet steekproeven gebruikt zijn. Het grootste deel van het internetpanel bestaat uit respondenten die geworven zijn via advertenties op websites. Wat betreft responspercentages van panel-respondenten is een respons van rond de 20% gebruikelijk, aldus Multiscope. Het is niet bekend waarmee de geconstateerde daling van het responspercentage, van 22,4% naar 13,4%, mee samenhangt.

De websites die betrokken waren bij het werven van respondenten bestonden voor onderzoeksjaar 2000 vooral uit zoekmachines en portals. Ondermeer de websites van Ilse, Vindex, Zoek en de Telefoongids werden gebruikt. In 2001 is een meer diverse verzameling van websites gebruikt voor de uitnodiging, van ondermeer Halfords, Hallmark, ANWB, Locatienet en 9292ov. Respons is aangemoedigd door het verloten van een prijs onder de respondenten, te weten drie boekenbonnen en een hoofdprijs in de vorm van reischeque. Het onderwerp van de enquête is niet vermeld bij de uitnodigingen, om een bias in de respons te voorkomen (Multiscope, 2001).

Het meest opvallende verschil tussen 2000 en 2001 vindt plaats in de samenstelling van de steekproef. In 2000 vormen panel-respondenten de grootste groep, terwijl die groep in 2001 in de minderheid is. De belangrijkste reden hiervoor heeft betrekking op grote groepen nieuwe internet-gebruikers die vooral in 2001 een enorme impuls hebben gegeven aan de groei van het internetgebruik. Het was met name de beschikbaarheid van zogenaamde “*gratis-internet*”-providers die deze grote groei mogelijk maakte. Omdat het internetpanel voornamelijk respondenten bevat die zich in de loop van de voorgaande jaren aangemeld hebben, waren in het internetpanel van Multiscope deze nieuwe groepen “*gratis-internet*”-gebruikers destijds ondervertegenwoordigd. Vandaar dat er in 2001 voor gekozen is om meer respondenten direct via websites te benaderen, zodat ook nieuwe internetgebruikers voldoende kans hadden om te participeren in het e-commerce onderzoek (Multiscope, 2003). Een ander belangrijk verschil tussen de werving van 2000 en 2001 heeft betrekking op de methodiek waarmee bezoekers benaderd werden. Bij de werving in 2000 zijn *banners* getoond aan bezoekers, terwijl in 2001 *pop-up windows* gebruikt zijn die verschenen aan elke 200ste bezoeker van de website. Deze stap van *banner* naar *pop-up* heeft met name betrekking op de eerder besproken effectiviteit van verschillende vormen van internet advertenties. Waar het responspercentage van *banners* al enige tijd aan het dalen was, en zich toendertijd onder de 1% begaf, was het responspercentage bij *pop-up windows*, die later op het internet geïntroduceerd werden dan *banners*, aanmerkelijk hoger. Een tweede voordeel van de *pop-up* methode is dat er ruimte is om een duidelijke uitnodiging te tonen, waardoor mensen de uitnodiging minder als advertentie beschouwen en eerder geneigd zullen zijn om mee te werken aan het onderzoek.

De enquêtegegevens zijn gewogen door middel van poststratificatie op een drietal variabelen, namelijk op basis van herkomst van de enquête, jaar van internettoegang en geslacht. Door op herkomst van de enquête te wegen, komt de verdeling van de reacties overeen met de werkelijke bezoekersaantallen over de verschillende websites.

Tot slot de relevante formuleringen met betrekking tot het onderzoeksonderwerp:

- *Internetgebruiker*: Een internetgebruiker is gedefinieerd als een persoon die tenminste één keer per week online is
- *Online buyer*: Nederlandse internetgebruikers die in het laatste jaar online producten of diensten aangekocht hebben, zowel voor privé als professioneel gebruik.

4.3 Resultaten

Voordat de verschillende resultaten naast elkaar gelegd kunnen worden, is het belangrijk om vast te stellen in hoeverre gegevens over dezelfde tijdsperiode verzameld zijn. In tabel 4.2 wordt een overzicht gegeven van de perioden waarin voor de verschillende onderzoeken enquêtes zijn afgenomen. Zowel NIPO als Multiscope hebben in het voorjaar gedurende één of twee maanden enquêtes afgenomen, zodat in hetzelfde najaar nog een onderzoeksrapportage over dat jaar gepubliceerd kon worden.

Zoals eerder vermeld verzamelt het CBS gegevens op een continue basis. Doordat er per maand cijfers met betrekking tot internetgebruik en elektronisch winkelen bekend zijn, is het mogelijk om de periode van dataverzameling in onze vergelijking zelf vast te stellen. Aangezien de andere twee onderzoeken enquêtes gebruiken die in het voorjaar afgenomen zijn, is ervoor gekozen om voor het CBS het gemiddelde te nemen van de maanden maart, april en mei, waardoor de perioden waarop resultaten betrekking hebben zo min mogelijk verschillen.

Tabel 4.2 Periode van dataverzameling per onderzoek

Jaar \ Onderzoek	Multiscope	NIPO	CBS
2000	mei juni	voorjaar	maart april mei
2001	april mei	februari	maart april mei

Bron: Multiscope (2001), NIPO (2001), CBS (2003)

Hoewel er wel verschillen zijn in de periode van dataverzameling, zijn deze echter vrij klein en mogen op basis daarvan geen grote verschillen in resultaten verwacht worden. De resultaten die achtereenvolgens besproken worden in deze paragraaf hebben respectievelijk betrekking op internetgebruik, online winkelen, populairste producten en redenen om niet online aankopen te doen. Op basis van de beschikbare gegevens was alleen bij online winkelen en populairste producten een vergelijking tussen alle drie de onderzoeksresultaten mogelijk, bij de overige vergelijkingen zijn twee onderzoeken betrokken. De resultaten van Multiscope verdienen bijzondere aandacht, omdat juist hier sprake is van online onderzoek met via het internet afgenomen enquêtes.

4.3.1 Internetgebruik

Alleen het CBS en NIPO hebben onderzocht hoe groot het percentage is van internetgebruikers onder de bevolking. Dat is ook logisch, aangezien Multiscope immers als onderzoekspopulatie “de internetgebruikers” heeft. Zoals ook aangegeven wordt in de verantwoording worden conclusies alleen getrokken over mensen (van 12 jaar of ouder) met internettoegang.

Wat betreft de resultaten, weergegeven in tabel 4.3, moet nog opgemerkt worden dat waar het CBS spreekt van *internettoegang*, het NIPO spreekt van *internetgebruik*. Er mag echter verondersteld worden dat deze twee percentages weinig van elkaar verschillen, het aantal mensen dat wel voorzieningen treft voor internet maar daar vervolgens geen gebruik van maakt, lijkt immers zeer beperkt. Multiscope (2003) merkt echter op dat deze verschillende begrippen wel degelijk kunnen leiden tot merkbaar verschillende percentages. Zo bestaan er in veel huishoudens gezinsleden die op zich wel toegang tot het internet hebben, maar er in de praktijk weinig tot geen gebruik van maken. Een dergelijk fenomeen zou een hoger percentage internettoegang bij het CBS voorspellen ten opzichte van het internetgebruik bij het NIPO-onderzoek, maar het tegendeel blijkt waar (tabel 4.3).

Tabel 4.3 Internet-toegang/gebruik, als percentage van de totale bevolking

Jaar \ Onderzoek	Multiscope	NIPO	CBS
2000	-	46	42
2001	-	52	54

Bron: Multiscope (2001), NIPO (2001), CBS (2003)

De twee onderzoeken, beide op basis van telefonische steekproeven, blijken niet veel te verschillen in hun resultaten wat betreft het internetgebruik. Waar in 2000 de CBS-schatting net iets lager was dan die van NIPO, schat CBS één jaar later het internetgebruik twee procentpunt hoger in dan NIPO. Bij dit laatste verschil kan het ook tijdsverschil in de dataverzameling, van twee maanden, een rol gespeeld hebben.

4.3.2 Online winkelen

Voor alle drie de onderzoeken is hiervoor een percentage bekend, wat gezien het onderzoeksonderwerp e-commerce ook niet verwonderlijk is. Dit percentage geeft immers het meest duidelijke antwoord op de simpele centrale vraag: “hoeveel mensen kopen online?”. De resultaten van alle drie de onderzoeksbureaus wat betreft het aantal online kopers als percentage van alle internetgebruikers, zijn weergegeven in tabel 4.4.

De resultaten in de CBS publicatie hadden als referentie de totale bevolking van Nederland. Om tot het aantal online kopers als percentage van alle *internetgebruikers* te komen, is voor het CBS het percentage personen van de Nederlandse bevolking die aan elektronisch winkelen doet (2000: 6% & 2001: 10%) gedeeld op het percentage personen met internettoegang, eerder weergegeven in tabel 4.3.

Tabel 4.4 Aantal online kopers, als percentage van alle internetgebruikers

Jaar \ Onderzoek	Multiscope	NIPO	CBS
2000	47	12	15
2001	32	18	19

Bron: Multiscope (2000/2001), NIPO (2001), CBS (2003)

We zien wederom dat het CBS en NIPO elkaar niet zo gek veel ontlopen. Aangezien zij beiden dezelfde methodiek gebruiken om een steekproef te trekken en telefonisch enquêtes afnemen, is dit ook niet zo verbazingwekkend. Het is onduidelijk waar de waargenomen verschillen, respectievelijk 3 en 1 procentpunt, aan toegeschreven kunnen worden.

Interessanter is het om te kijken of Multiscope, die haar cijfers baseert op steekproeven onder internetgebruikers zelf, met vergelijkbare resultaten komt. Het blijkt dat Multiscope op veel hogere percentages uitkomt dan haar concurrenten. Ook blijkt bij Multiscope echter een opvallende daling ingezet te hebben in het percentage mensen die als “*online buyers*” te boek staan tussen 2000 en 2001. Deze merkwaardige trend behoeft nadere toelichting.

De rapportage van Multiscope (2001) geeft weliswaar geen directe verklaring hiervoor, maar merkt wel op dat in haar voorlaatste rapportage over 2000, gepubliceerd vóór de stagnatie van de internetsector als geheel (na het eind van de internethype, maart 2000), wellicht een te optimistisch beeld geschetst was. Verdere verklaringen van Multiscope (2003) voor deze daling hebben met name betrekking op de opkomst van het gratis-internet, waardoor een grote groep nieuwe internetgebruikers aangetrokken werd tot het (gratis) internet. Nieuwe internetgebruikers blijken gedurende de eerste 2 jaar van hun internetgebruik nauwelijks aankopen online te doen, en zodoende zou dit een negatieve invloed kunnen hebben op het percentage online kopers aangezien het totaal aantal internetgebruikers wel toeneemt terwijl het aantal online kopers niet toeneemt.

Toch kunnen de deelverklaringen, betrekking hebbende op “gratis internet”-gebruikers en een gedaald vertrouwen van internetgebruikers, naar de mening van de schrijver van deze scriptie niet verklaren waarom een dergelijk sterke daling zich voorgedaan heeft, terwijl deze daling niet zichtbaar is bij de resultaten van de andere twee onderzoeken van het CBS en NIPO. Als vervolgens gekeken wordt naar de steekproef samenstelling van Multiscope in 2000 en 2001, kan een overschatting van het percentage online kopers in met name 2000 mogelijk veroorzaakt worden door het relatief grotere aantal panel-respondenten in 2000 (waarin online kopers oververtegenwoordigd zouden zijn), of een relatief grotere *self-selecting* bias bij werving met *banners* ten opzichte van werving middels *pop-up windows*. Aangetekend zij dat dit suggesties zijn die weliswaar overeenkomen met eerdere bevindingen in deze scriptie, maar die hier verder niet onderbouwd of bewezen kunnen worden door middel van statistische analyse.

Afgezien van deze daling, liggen de absolute waarden van de percentages van Multiscope ook allebei ver boven die van het CBS en NIPO, die traditionele en beproefde telefonische enquêtes als vooralsnog meer betrouwbaar uitgangspunt nemen. Multiscope (2003) geeft met betrekking tot dit verschil aan dat definitieverschillen hiermee verband kunnen houden. Zo rekent Multiscope respondenten die aankopen doen voor zakelijk gebruik ook tot de *online buyers* groep, in tegenstelling tot beide andere onderzoeken. Als gekeken wordt naar de laatste aankoop volgens de Multiscope rapportage, blijkt inderdaad 7% betrekking te hebben op professioneel gebruik tegenover 93% voor privé gebruik. Dit verschil, en andere niet verder te specificeren definitieverschillen, kunnen echter niet de forse omvang van de geconstateerde verschillen in percentage online kopers verklaren.

Als gekeken wordt naar de wervingsmethoden die aan de basis staan van het Multiscope onderzoek, lijkt een invloed van de *self-selecting* aard van de online werving niet onwaarschijnlijk. Het zouden dan met name respondenten zijn die relatief vaak ervaring hebben met kopen op het internet, die door het proces van *self-selection* sterk oververtegenwoordigd zijn in de steekproef, zowel bij het internetpanel als bij de werving op websites. Weging van gegevens, op jaar van internettoegang en andere demografische variabelen, heeft hier dan niet voor kunnen corrigeren.

4.3.3 Meeste online gekochte producten

Alle onderzoeken laten tevens zien welke producten en diensten het meest populair zijn onder de online kopers. De vraagstelling in de enquêtes verschilt echter per onderzoek. Zo heeft Multiscope gevraagd naar de categorie waarin de laatste aankoop van de respondent viel. Bij het NIPO onderzoek werd gevraagd naar producten die de respondent online had gekocht of besteld, hierop waren meerdere antwoorden mogelijk. Omdat het totaal percentage hiermee boven de honderd procent uitkwam en dus niet geschikt was voor vergelijking, is de verdeling van antwoorden herleid tot een totaal van honderd, zodat vergelijking beter mogelijk is. De uitkomsten van het CBS tot slot hebben ten eerste betrekking op het najaar van 2001 (november), terwijl de waarden het percentage van aankopen weergeven van goederen en diensten die in de drie maanden voorafgaand aan de enquête gekocht zijn. Daarnaast moesten enkele categorieën samengevoegd worden, voordat de indeling van de antwoordcategorieën min of meer overeen kwam.

Desalniettemin laat tabel 4.5 zien dat de resultaten verrassend dicht bij elkaar liggen.

Tabel 4.5 Top zeven online gekochte producten/diensten, in procenten (2001)

Product \ Onderzoek	Multiscope	NIPO	CBS ¹
Boeken	20	21	22
Muziek CD's	15	18	16 ²
Reizen	10	4	9
Kleding	10	9	12
PC Software	7	8	3
Hardware	10	9	9
Video / DVD	5	5	n.b. ²
Overig	28	29	22
Totaal	100	100	100

Bron: Multiscope (2001), NIPO (2001), CBS (2003)

¹) November 2001

²) CBS kent 16 procent voor totaal van muziek CD's, Video & DVD

En in tegenstelling tot de vorige vergelijking, blijkt de positie van Multiscope als online onderzoeker zeer vergelijkbare resultaten te geven. Op de categorie muziek CD's ná, heeft Multiscope zelfs bij alle categorieën een verschil van 1 procentpunt of minder ten opzichte van het CBS of NIPO. Dit geldt andersom niet, waarmee Multiscope zich in het betrouwbare midden lijkt te positioneren ten opzichte van haar concurrenten, als het gaat om de meeste online gekochte producten of diensten. Wellicht kunnen vragen met betrekking tot online aankoopgedrag beter geënuquêteerd worden met behulp van het internet in plaats van telefonisch. Een *self-selecting* bias die de resultaten negatief beïnvloedt, lijkt hier niet aanwezig te zijn.

4.3.4 Redenen om niet online te kopen

Als laatste worden twee vergelijkbare resultaten bekeken van Multiscope en NIPO, die betrekking hebben op de redenen die respondenten aanvoeren op de vraag waarom ze (nog) niet online producten of diensten kopen. Het gaat hier dus om respondenten die al eerder in de enquête aangegeven hebben niet tot de online kopers-groep te behoren.

Bij beide onderzoeken waren meerdere antwoorden mogelijk, dus totalen tellen niet op tot 100%. Bij dergelijke vragen is het natuurlijk de indeling van de antwoordcategorieën, zoals die bij het opstellen van de enquête bedacht wordt, die een sterke invloed heeft op de resultaten. Maar hoewel er verschillende formuleringen voor antwoorden gebruikt worden, komt de strekking van de verschillende antwoorden in het algemeen toch wel overeen, zo laat tabel 4.6 ook zien.

Tabel 4.6 Redenen om niet online aankopen te doen (2001)

Multiscope		NIPO	
Gebrek aan vertrouwen in online betaling	46%	45%	Het is makkelijker/leuker om producten/diensten in de winkel te kopen
Koop liever offline	39%	42%	Wil mijn creditcard nummer niet geven / bang voor veiligheidsproblemen
Bezit geen creditcard	36%	35%	Het is veiliger om producten/diensten in de winkel te kopen
Ontbreken persoonlijk contact verkoper/adviseur	22%	29%	Je weet niet wat je krijgt
Gebrek aan vertrouwen in internet	19%	16%	Geen vertrouwen in online merken (gebrek aan betrouwbaarheid)
Nog te onervaren / onbekend met internet	18%	8%	Te duur, verwacht lager prijzen op internet
Gebrek aan vertrouwen in de aanbieders	15%	7%	Te moeilijk / weet niet hoe dat werkt
De gegeven informatie over product / aanbieder is te beperkt	9%	7%	De producten en diensten op internet zijn niet interessant genoeg
Te beperkt aanbod	6%	6%	Levertijd is te lang en andere leveringsproblemen
Weet niet hoe het moet	5%	1%	Creditcard maatschappij gaf geen goedkeuring/niet voldoende saldo
Oninteressant aanbod	5%	33%	Anders
Kan het gewenste / gezochte product niet vinden	4%		
Vind het te duur	4%		

Bron: Multiscope (2001), NIPO (2001)

Hoewel verschillende antwoordcategorieën een directe vergelijking niet mogelijk maken, komen de belangrijkste redenen om niet online aankopen te doen redelijk goed overeen. Zo heeft de top drie van antwoorden bij beide onderzoeken betrekking op dezelfde redenen, namelijk veiligheid online betalen (Multiscope:1 & NIPO:3), voorkeur offline winkelen (Multiscope:2 & NIPO:1) en creditcard problemen (Multiscope:3 en NIPO:2).

De overige redenen laten iets grotere verschillen in absolute percentages zien, maar dat heeft waarschijnlijk te maken met de formuleringen van verschillende antwoordcategorieën en de manier van enquêteren (via computerscherm vs. telefonisch). Online enquêtes kunnen bij een dergelijke vraag wellicht tot meer betrouwbare resultaten leiden, omdat de respondent direct

een overzicht heeft op zijn scherm van alle mogelijke antwoordcategorieën. Bij telefonisch onderzoek wordt een groter beroep gedaan op het geheugen en inbeeldingsvermogen van de respondenten, waardoor een gebrek aan overzicht tot onjuiste antwoorden kan leiden.

Interview Multiscope

Om de resultaten van dit vergelijkende hoofdstuk enigszins in perspectief te plaatsen, wordt in een interview John Kivit, directeur van Multiscope, om een reactie gevraagd. Hieronder volgt een korte samenvatting van dit interview met betrekking tot de hier besproken *self-selecting* bias en de toekomst van online onderzoek.

Bias

Teveel nadruk wordt in de ogen van Multiscope gelegd op de *self-selecting* bias, waarvan de omvang en het belang nog niet bewezen is. Met betrekking tot het percentages online kopers, wordt een invloed van een self-selecting bias niet uitgesloten maar wordt tevens aangegeven het niet waarschijnlijk lijkt dat deze meer dan 5 procentpunt bedraagt in het uiteindelijke percentage online kopers. Door het gebruik van pre- en post-stratificatie wordt bij Multiscope op de belangrijkste demografische variabelen (inbegrepen specifieke variabelen over de internetsituatie) voor een deel van deze bias gecorrigeerd. Er wordt op gewezen dat ook in telefonisch onderzoek een bias bestaat door de selectiviteit van non-respons. Zo waren “heavy internetters” (>15 uur per week) tot een paar jaar geleden bijvoorbeeld ondervertegenwoordigd in RDD-telefonische onderzoeken, simpelweg omdat ze vaak in gesprek waren – de internetverbinding hield de telefoonlijn bezet. Met betrekking tot de verschillen in het percentages online kopers zijn er volgens Multiscope andere factoren, met name definitieverschillen, die een grotere bijdrage leveren aan de geconstateerde verschillen.

Toekomst online onderzoek.

Over het algemeen is de verwachting dat respons percentages over de hele linie zullen blijven dalen, en mensen steeds meer de behoefte en de (technische) mogelijkheden krijgen om contacten te filteren op de wenselijkheid van die contacten. Mensen zijn in toenemende mate “onderzoeksmoe”, en zeker Nederland is wat dat betreft een “overzocht” land.

Een consequentie van dalende responspercentages kan zijn dat het belonen van respons meer gemeengoed gaat worden, zelfs als dat betekent dat er een groter risico op selectieve respons bestaat. Anderzijds kan onderzoek in de toekomst ook steeds meer plaatsvinden op basis van directe registraties van gedrag in elektronische databestanden. Het analyseren van grote gegevens bestanden, zoals bijvoorbeeld Interpay die heeft over aankoopgedrag en internet-service providers over surfgedrag, staat ook wel bekend als *data-mining*. De verdere automatisering van allerlei dagelijkse handelingen en processen, die gepaard gaat met vergaande registratie van gegevens, kan er voor zorgen dat deze vorm van onderzoek aan belang gaat winnen.

4.4 Conclusie

Zowel het CBS als het NIPO onderzoek zijn uitgevoerd met gebruikmaking van traditionele onderzoeksmethoden, dat wil zeggen de gegevens zijn verkregen door middel van een willekeurige steekproeftrekking en telefonische enquêtes. Het Multiscope onderzoek heeft zowel bij de werving als bij het afnemen van de enquêtes gebruik gemaakt van internet, en valt zodoende onder de noemer “online onderzoek”.

De kwesties waarop de e-commerce onderzoeken meer inzicht wilden bieden, hadden als eerste betrekking op de vraag hoeveel procent van de internetgebruikers online aankopen deed, als tweede op de kenmerken van de online kopers zelf, en als derde op de kenmerken van de overige groep internetgebruikers (niet-online kopers).

De eerste vraag wordt doorgaans beantwoord door gevonden verhoudingen in de steekproef te generaliseren naar de totale internetpopulatie. De vergelijking van de drie onderzoeken op dit punt (paragraaf 4.3.2), laat zien dat online onderzoek waarbij gebruik wordt gemaakt van *self-selecting* wervingsmethoden, niet geschikt is gebleken voor het maken van generalisaties naar een grotere populatie, zelfs als die grotere populatie nog beperkt is tot internetgebruikers.

Wat betreft de tweede deelvraag naar de kenmerken van online kopers, is in paragraaf 4.3.3 gebleken dat waar het gaat om categoriale schattingen, het gebruiken van online onderzoek zeer geschikt kan zijn. Wellicht draagt de oververtegenwoordiging van online kopers bij aan de betrouwbaarheid van de schattingen, omdat de vraag juist alleen van toepassing was op deze oververtegenwoordigde doelgroep van online kopers.

Bij de derde deelvraag naar de kenmerken van de overige groep internetgebruikers, zijn redenen die respondenten aanvoeren op de vraag waarom niet online gekocht wordt in kaart gebracht. Ook hier lijkt online onderzoek in staat tot het maken van redelijke schattingen van op categorieën gebaseerde resultaten. Omdat hier alleen een vergelijking met één traditioneel (NIPO) onderzoek mogelijk was, is het echter niet duidelijk in hoeverre geconstateerde verschillen het gevolg zijn van de gehanteerde methode (traditioneel vs. online onderzoek) of van andere factoren.

Resumerend is gebleken dat online onderzoek goed geschikt kan zijn voor het schatten van categoriale variabelen, in het bijzonder waar dat gebeurt voor een specifieke doelgroep zoals hier het geval was bij “online kopers”. Voor het maken van absolute generalisaties lijkt online onderzoek, gebaseerd op een *self-selecting* wervingsmethode, echter minder geschikt. De resultaten van de hier uitgevoerde vergelijking tussen drie e-commerce onderzoeken lijken de in paragraaf 3.3.2 gemaakte beweringen te ondersteunen.

5 Conclusie

Met deze scriptie is getracht meer inzicht te krijgen in de wereld van online onderzoek. De betrouwbaarheid van online onderzoeksresultaten stond daarbij centraal. In de inleiding werd de volgende probleemstelling gegeven:

"In hoeverre kunnen door middel van online onderzoek kwalitatief goede, betrouwbare en representatieve gegevens verkregen worden, en in welke mate moet de methodologische aanpak van online onderzoek verschillen van de traditionele onderzoeksmethodiek."

Wat betreft het tweede gedeelte van deze probleemstelling is duidelijk geworden dat online onderzoek een geheel nieuwe discipline binnen kwantitatief onderzoek is, die veel nieuwe eigenschappen, kenmerken en mogelijke problemen met zich mee brengt die niet van toepassing zijn op traditionele onderzoeksmethoden. Zo zijn er bijvoorbeeld nieuwe mogelijkheden wat betreft de presentatie en het ontwerp van enquêtes en de communicatie met respondenten, die indien ze juist gebruikt worden kunnen bijdragen tot een betere kwaliteit van de te verzamelen enquêtegegevens. Ook de ontwikkeling en het relatieve succes van internetpanels laat zien dat online onderzoek om een nieuwe methodiek en benadering vraagt.

Het is moeilijker om over het eerste deel van de vraag, de hoofdvraag over de betrouwbaarheid van online onderzoek, eenduidige conclusies te trekken. Wordt gekeken naar ervaringen in het recente verleden met online onderzoek, dan moet geconstateerd worden dat tot nu toe de betrouwbaarheid en representativiteit van traditioneel onderzoek nog beter is dan die van online onderzoek. Het beeld dat naar voren komt in diverse publicaties is wel dat online onderzoek bezig is met het maken van een inhaalslag.

Er bestaan echter een aantal knelpunten, met als belangrijkste het ontbreken van een methodiek om een willekeurige steekproef te trekken op het internet, die betrouwbare en generaliseerbare online onderzoeksresultaten in de weg staan. Immers is de hele statistische kwantitatieve wetenschap gebaseerd op die ene belangrijke vooronderstelling, namelijk dat iedereen in de populatie een gelijke (of bekende) kans heeft om in de steekproef getrokken te worden. Zolang niet aan die voorwaarde voldaan kan worden, lijkt het moeilijk dat online onderzoek de betrouwbaarheid die geassocieerd wordt met RDD-telefonische steekproeven kan evenaren.

Opgemerkt zij wel dat dit probleem vooral speelt bij vormen van onderzoek waarbij generalisaties naar een grotere, vaak landelijke, populatie gemaakt moeten worden. Met name internetpanels kunnen een belangrijke en nuttige rol vervullen op andere gebieden, zoals het in kaart brengen van ontwikkelingen, categoriale variabelen en doelgroep onderzoek. Dit wordt in deze scriptie nogmaals bevestigd door de uitgevoerde casestudy waarin een online onderzoek vergeleken wordt met twee traditioneel uitgevoerde onderzoeken.

Omdat zowel online onderzoek als tal van andere internetgerelateerde fenomenen een snelle ontwikkeling doormaken, is het moeilijk om online onderzoek vanuit een methodologische invalshoek bij te houden. Zo werd bijvoorbeeld tien jaar geleden geconstateerd dat door middel van e-mail het makkelijk was om respondenten te werven, maar inmiddels heeft een ander fenomeen, namelijk *spam*, roet in het eten gegooid. Om een geaccepteerde wetenschappelijk verantwoorde methodologische basis voor online onderzoek ontwikkeld te zien is veel meer tijd en onderzoek nodig.

Afsluitend beseft de schrijver dezes zich, dat scepsis in deze scriptie wellicht een te grote rol toebedeeld is. Immers heeft de ervaring geleerd dat ontwikkelingen moeilijk te voorspellen zijn, zeker op het vlak van informatie- en communicatietechnologie. Ook bij online onderzoek kunnen nieuwe ontwikkelingen misschien mogelijkheden scheppen tot het trekken van willekeurige steekproeven online, of de noodzaak van een willekeurig getrokken steekproef helemaal overbodig maken. Voor de visionair die hierbij kan voorspellen welke methode succesvol zal blijken, is een gouden toekomst weggelegd.

Literatuurlijst

Alvarez, R.M. & VanBeselaere, C. (2002), "Web-based surveys", Article prepared for the Encyclopedia of Measurement, <http://survey.caltech.edu/encyclopedia.pdf>

Bethlehem, J. (1997), "*Bacula : Current Status and Future Developments*", Centraal Bureau voor de Statistiek, <http://www.blaiseusers.org/Ibucpdfs/1995-1998/bethle97.pdf>

Blyth F.M., March L.M., Shellard D. & Cousins M.J. (2002), "*The experience of using random digit dialling methods in a population-based chronic pain study*", Australia and New Zealand Journal of Public Health, Vol. 26-6, December 2002, p. 511-514.

CBS (2002), "*De Digitale Economie 2002*", Centraal Bureau voor de Statistiek, Voorburg/Heerlen, <http://www.cbs.nl/nl/publicaties/publicaties/bedrijfsleven/algemeen/p34-02.pdf>

CBS (2003), "*De Digitale Economie 2003*", Centraal Bureau voor de Statistiek, Voorburg/Heerlen, <http://www.cbs.nl/nl/publicaties/publicaties/bedrijfsleven/algemeen/p34-03.pdf>

Cook, C., Heath, F. & Thompson, R. (2000), "*A meta-analysis of response rates in web or Internet-based surveys*", Educational and Psychological Measurement Vol. 60, p. 821-836.

Cooper, K. & Victory, N. (2002), "*A Nation Online: How Americans are Expanding Their Use of the Internet*", National Telecommunications and Information Administration and Economics and Statistics Administration, Washington, <http://www.ntia.doc.gov/ntiahome/>

Couper, M.P. (2002), "*New Technologies and Survey Data Collection: Challenges and Opportunities*", Paper presented at the International Conference on Improving Surveys, August 25-28 2002, Copenhagen, http://www.icis.dk/ICIS_papers/

Crawford, S., McCabe, S., Couper, M. & Boyd, C. (2002), "*From Mail to Web: Improving Response Rates and Data Collection Efficiencies*", paper presented at the International Conference on Improving Surveys 2002, August 25-28 2002, Copenhagen, http://ww.icis.dk/ICIS_papers/

Danielsson, S. (2002), "*The propensity score and estimation in nonrandom surveys – an overview*", Modern statistical survey methods project report no. 18, Dept. of Statistics, University of Linköping, Sweden.

Dennis, J.M. (2003), "*An Overview of Capabilities and Methodological Research Conducted by the Government and Academic Area, Knowledge Networks*", Knowledge Networks, Menlo Park (VS).

Dillman, D.A. (1998), "*Mail and Other Self-Administered Surveys in the 21st Century: The Beginning of a New Era*", Social Economic Sciences Research Center, Washington State University, Washington, <http://survey.sesrc.wsu.edu/dillman/papers/>

- Dillman, D.A. (1978), *“Mail and Telephone Surveys: The Total Design Method”*, Wiley-Interscience, New York.
- Dillman, D.A. (2000), *“Mail and Internet Surveys, The Tailored Design Method”*, John Wiley & Sons, New York.
- Dillman, D.A. & Carley-Baxter, L.R. (2000), *“Structural Determinants of Mail Survey Response Rates Over a 12 Year Period, 1988-1999”*, Proceedings of the section on survey method, The American Statistical Association, Alexandria (VA), <http://survey.sesrc.wsu.edu/dillman/papers/>
- Dillman, D.A., & Bowker, D.K. (2001), *“The Web Questionnaire Challenge to Survey Methodologists”*. In Bernad Batinic, Ulf-Dietrich Reips, Michael Bosnjak & Andreas Werner: Online Social Sciences, p. 53-71, Hogrefe & Huber, Seattle.
- Dreze, X., Kalyanam, K. & Briggs, R. (2002), *“Increasing Panel Data Accuracy: An Application to Internet Panels”*, Work in Progress, as published on website <http://www.xdreze.org/Publications/>
- Emerce (2001), *“Een kwestie van definities”*, Emerce Business Innovation Media & Technologie nr. 16, Maart 2001, <http://www.emerce.nl/archives/magazine/maart2001/update/12146.html>
- Feld, K.G. (2001), *“Online polling and survey research”*, Campaigns & Elections Vol. 22, p. 41-43.
- Foxnews (2003), *“Telemarketers Sue Over Do-Not-Call List”*, <http://www.foxnews.com/story/0,2933,93091,00.html>
- Gouw, Mariola de & Groenland, Edward. (2000), *“Propensity Weighting: een nieuwe methode voor het wegen van onderzoeksgegevens”*, Blauw Research, Rotterdam, <http://www.blauw.com/docs/docs/pw-download.pdf>
- Grossnickle, Joshua & Raskin, Oliver (1999), *“Market Research on the Web”*, <http://hotwired.lycos.com/webmonkey/e-business/marketing/tutorials/tutorial4.html>
- Harris Interactive (2003), *“An Interview With Our Panel Expert & Methodology Expert”*, <http://www.harrisinteractive.com>
- Heerwegh D. & Loosveldt, G. (2002), *“Radio-buttons or select menus? An evaluation of the effect of response formats on data quality in Web surveys”*, Paper presented at the International Conference on Improving Surveys (ICIS), August 25-28 2002, Copenhagen.
- Kim, K. (2003), *“Suggestions to improve data quality: Resolving problems in online data collection”*, Quantitative Methods in Advertising and Brand Communication, Center for Interactive Advertising (CIA), University of Texas, Austin, <http://www.ciadvertising.org>
- Krotki, K. & Dennis, J.M. (2001), *“Probability-Based Survey Research on the Internet”*, Paper to be presented at the 53rd Conference of the International Statistical Institute, August 22-29th 2001, Seoul, <http://www.rti-knowledgenetworks.com/isi-2001-conf-paper.pdf>

Lynch, B.P. (2002), "*The Digital Divide or the Digital Connection: a U.S. Perspective*", A paper presented to the Scottish Library Association, May 20th 2002, Peebles, Scotland, <http://www.slainte.org.uk/CILIPS/BGDay2002/Lynchb.pdf>

MacElroy, B., Milucki, J., and Mc Dowell, P. (2002), "*A Comparison of Quality in Open-end Responses and Response Rates Between Web-based and Paper and Pencil Survey Modes*", Journal of Online Research, <http://www.ijor.org>

Manfreda, K.L. & Vehovar, V. (2002), "*Survey design features influencing response rates in web surveys*", Paper presented at the ICIS 2002 - International Conference on Improving Surveys, August 25-28th 2002, Copenhagen, http://www.icis.dk/ICIS_papers/C2_4_3.pdf

Miller, Thomas.W. and Panjikaran, Kurian.J. (2001), "*Studies in Comparability: The Propensity Scoring Approach*", A.C. Nielsen Center for Marketing Research, University of Wisconsin-Madison, Madison.

Molenaar, Cor (1999), "*Veranderingen door Internet, Branches in de toekomst*", Samson Publishing, Alphen aan de Rijn.

Molenaar, Cor (2001), "*E-strategie, nieuwe regels voor de oude economie*", Pearson Education, Amsterdam.

Multiscope (2000), "*E-commerce – the Netherlands, 2000*", Multiscope, Den Bosch.

Multiscope (2001), "*E-commerce – the Netherlands, 2001*", Multiscope, Den Bosch.

Multiscope (2003), Interview met John Kivit (Directeur Multiscope), Juli 2003 & November 2003, Den Bosch.

Newsweek (1991), "*Newsweek Media Research Index: General Research - Methodological Research*", <http://www.vmr.com/research/910.html>

Nielsen//Netratings (2003), "*Global Internet Population Grows An Average of Four Percent Year-Over-Year*", Nielsen//Netratings Press Release, april 2003, http://www.nielsen-netratings.com/pr/pr_030220_hk.pdf

Nielsen Media Australia (2003), "*580 million people now have internet access*", Nielsen Media Australia Website, <http://www.nielsenmedia.com.au/industry.asp?industryID=22>

NIPO (2001), "*Wereldwijd E-Commerce Onderzoek 2001*", NIPO Interactive, Amsterdam.

Nyffjäll, Mats. (2001), "*Can Internet panels be used for generalizations to the general public? A study based on two parallel surveys in Sweden*", Statisticon Statistics & Research, Uppsala, Sweden, <http://www.netsurvey.se/offices/Sweden/pdf/Report%20parallel%20study2.pdf>

Rizzi, John (2002), "*E-Mail Is Growing Up - Fast*", Press Release e-Dialog, http://www.e-dialog.com/press_20020109.shtml

Schonlau, M., Fricker, R.D. Jr. & Elliott, M.N. (2001), "*Conducting Research Surveys via E-mail and the Web*", RAND, Santa Monica, California, <http://www.rand.org/publications/MR/MR1480/>

Schonlau, M., Zapert, K., Simon, L.P., Sanstad, K., Marcus, S., Adams, J., Sprance, M., Kan, H., Turner, R. & Berry, S. (2003), "*A Comparison Between Responses From a Propensity-Weighted Web Survey and a identical RDD Survey*", *Social Science Computer Review* 21, http://www.schonlau.net/publication/03socialsciencecomputerreview_propensity_galley.pdf

Sheehan, K.B. (2002), "*Online Research Methodology: Reflections and Speculations*", *Journal of Interactive Advertising* Vol. 3, Nr. 1 2002, <http://jiad.org/vol3/no1/sheehan/>

Sikkel, Dirk (2000), "*CentERpanel, representativiteit en weging*", Stichting CentERdata, Sociaal Cultureel Planbureau, Rijswijk, <http://www.scp.nl/boeken/titels/2003-1/nl/acrobat/wegen.pdf>

Smith, T.W. (2001), "*Are representative internet surveys possible?*", Proceedings presented at the Statistics Canada Symposium 2001, <http://www.statcan.ca/english/conferences/symposium2002/sessionsproceed.htm>

Spit, Jan (1996), "*Onderzoeksverantwoording Aanvullend Voorzieningengebruik Onderzoek 1995*", Sociaal en Cultureel Planbureau, Rijswijk, <http://www.scp.nl/boeken/werkdocumenten/doc37/nl/acrobat/docu37.pdf>

Talyor, Humprey & Terhanian, George (1999), "*Heady Days Are Here Again, Online polling is rapidly coming of age*", *Public Perspective* June/July 1999, p. 20-23, <http://www.ropercenter.uconn.edu/pubper/pdf/pp104b.pdf>

TNO-STB (1999), "*Bouwstenen voor een Digitaal trapveld*", TNO-STB in opdracht van Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, oktober 1999, http://www.stadsadviseur.nl/handreikingen/pdf/bouwstenen_digitaal_trapveld.pdf

Topp, Neal W. & Pawloski, Bob (2002), "*Online Data Collection*", *Journal of Science Education and Technology*, Vol. 11 No. 2, Plenum/Kluwer Publishing.

Totten, Jeff. W. (2002), "*Use of e-mail and internet surveys by research companies*", *Journal of Online Research*, <http://www.ijor.org>

Varedian, M. & Forsman, G. (2003), "*Comparing Propensity Score Weighting with Other Weighting Methods: A Case Study on Web Data*", Department of Statistics at Linköpings Universitet, Sweden, <http://www.zapera.com/images/news/2003/ID109/fig1.PDF>

Veen, F.V. van der (2003), "*Determinantenonderzoek: PC-bezig, internettoegang en elektronisch winkelen onder de Nederlandse bevolking*", Centraal Bureau voor de Statistiek, Voorburg/Heerlen.

Vehovar, V. & Manfreda, K. L. (1999), "*Web Surveys: Can the Weighting Solve the Problem?*", 1999 Proceedings of the Section on Survey Research Methods, American Statistical Association, http://www.amstat.org/sections/srms/Proceedings/papers/1999_168.pdf

Volkskrant (2003), "*Bel-me-niet-lijst VS onwettig verklaard*", Volkskrant 26 september 2003.

Webwereld (2003), "*Webonderzoeker Pro Active failliet verklaard*", Webwereld.nl editie 18 april 2003, <http://www.webwereld.nl/nieuws/14821.phtml>

Willemsen, A. & Lingeman, H. (2002), "*Elixer voor dalende response; mixed mode onderzoek*", <http://www.interview-nss.com/mixed.pdf>