

HANDLEIDING DATAVERZAMELING

met behulp van randomized response

Capaciteitsgroep Methodenleer & Statistiek
Universiteit Utrecht

Voorwoord

Het Expertisecentrum Rechtshandhaving (ERh) van het Ministerie van Justitie begeleidt en ontwikkelt onderzoek naar de mate van naleving van instrumentele wetgeving. Ten behoeve van de dataverzameling voor een dergelijke nalevingsmonitor bestaat een uitgebreid onderzoeksinstrumentarium waarbinnen de enquête van de doelgroep een centrale plaats inneemt. Gezien het sensitieve karakter is ervoor gekozen deze enquêtes volgens de randomized responsemethodiek uit te voeren, deze methodiek is speciaal ontwikkeld voor het verzamelen van sensitieve data. Randomized response is echter niet de enige methode voor het verzamelen van sensitieve data en binnen de randomized response bestaan er diverse varianten.

In het kader van de begeleiding van zowel opdrachtgevers als uitvoerders van nalevingsmonitor heeft het ERh de capaciteitsgroep Methodenleer & Statistiek van de Universiteit Utrecht verzocht een handleiding te maken met een overzicht van de meest gangbare methoden voor het verzamelen van sensitieve data en een instructie voor het uitvoeren van een randomized response-enquête volgens de ‘current best’ methode. Dit heeft geresulteerd in de ”Handleiding Dataverzameling met behulp van Randomized Response”. In hetzelfde kader is ook de “Handleiding Data-analyse bij gebruik van Randomized Response” verschenen, waarin wordt ingegaan op statistische technieken voor het analyseren van data die met randomized response zijn verzameld.

September 2003

Inhoudsopgave

Voorwoord	iii
Samenvatting	v
Inleiding	5
1. Computergestuurde randomized response (CARR)	7
1.2 Instructie vragenlijst	8
1.3 Toelichting op de instructies	11
2. Alternatieve technieken binnen randomized response	13
2.1 Randomized Responsetechnieken	13
2.1.1 Warner's originele methode	14
2.1.2 De forced-responsemethode	15
2.1.3 Unrelated Question design (UQD)	17
2.1.4 Kuk's kaartmethode	19
2.2 Relatieve efficiëntie en steekproefgrootte	20
2.3 Betrouwbaarheidsinterval en steekproefgrootte	22
2.4 Steekproefgrootte in eindige populaties	23
3. Alternatieven voor randomized responsetechnieken	25
3.1 Het computer gestuurde interview (CASI)	25
3.2 'Locked box'-design	26
3.3 De unmatched count techniek	27
4. Verder lezen	29
4.1 Nederlands	29
4.2 Engels reviews	29
4.3 Engels methodebeschrijving	30
4.4 Engels algemeen vergelijkend onderzoek van 1990-2003	31
Bijlage A: Betrouwbaarheidsintervallen oneindige populaties	33
Bijlage B: Betrouwbaarheidsintervallen eindige populaties	37

Samenvatting

Van de methoden voor het verzamelen van sensitieve data is de computergestuurde zelfrapportage met randomized response (CARR) volgens de huidige inzichten het beste. De handleiding laat aan de hand van een praktijkvoorbeeld van een onderzoek naar de mate van naleving door uitkeringsgerechtigden zien hoe deze methode in de praktijk kan worden toegepast. De CARR-methode is makkelijk uit te leggen, efficiënt en valide in de zin dat de respondenten geneigd zijn tot grotere openheid.

Binnen randomized response bestaan een aantal varianten, waarvan er vier in deze handleiding worden besproken. Van deze methoden wordt *forced response* momenteel als de beste gezien vanwege de combinatie van een hoge efficiëntie en valide onderzoeksresultaten. Daardoor kan - vergeleken met de andere randomized responsevarianten - met aanzienlijk kleinere steekproeven worden volstaan.

De aanbevolen steekproefomvang voor een onderzoek dat gebruik maakt van forced response bedraagt circa 750. Bij de berekening van dit aantal is van een oneindig grote populatie uitgegaan en een percentage overtreders van 7% in de steekproef, het percentage overtreders in de populatie ligt dan met 95% zekerheid tussen de 3% en de 11%. Bij kleinere steekproeven neemt de omvang van dit betrouwbaarheidsinterval toe, en worden de schattingen al snel te onnauwkeurig. In twee situaties kan worden afgeweken van deze aanbevolen minimale steekproefomvang. Wanneer op voorhand kan worden verwacht dat het percentage overtreders veel hoger is dan 7% en ongeveer tussen de 25% en 75% ligt kan in principe met een kleinere steekproef worden volstaan. De tabel in bijlage A kan gebruikt worden om de gewenste steekproefomvang te bepalen. De tweede uitzondering vormt de situatie dat de populatie klein van omvang is. Met de tabel in bijlage B kan de gewenste steekproefomvang voor eindige populaties bepaald worden.

De meest gangbare alternatieven voor randomized response zijn de *'locked box'-methode*, de *unmatched count techniek* en *computergestuurde zelfrapportage* (CASI). De eerste twee methoden zijn ofwel wat betreft validiteit ofwel wat betreft efficiëntie inferieur aan de forced-responsemethode. De CASI-methode is qua validiteit en efficiëntie bevredigend en is met de forced-responsetechniek te combineren tot de CARR-methode.

Inleiding

In het algemeen kunnen onderzoeksvragen op twee manieren bedreigend voor respondenten zijn, ze kunnen een *intrinsieke* of een *extrinsieke* dreiging vormen (Lee, 1993). Er is sprake van een *intrinsieke dreiging* wanneer de vragen betrekking hebben op onderwerpen die zeer persoonlijk van aard zijn, pijnlijk zijn, of stressvol. De dreiging komt dan vanuit de respondent zelf. Er is sprake van *extrinsieke dreiging* als de antwoorden sancties kunnen oproepen. In het geval van regelovertreding bijvoorbeeld kunnen de intrinsieke en extrinsieke dreiging gevormd worden door respectievelijk de gêne die de respondent voelt om de overtreiding toe te geven en de angst van de respondent voor mogelijke strafrechtelijke sancties wanneer de antwoorden bekend worden en worden doorgegeven aan officiële instanties. Om deze redenen worden regelovertredingen dan ook vaak ondergerapporteerd.

De onderrapportage van sensitief gedrag is problematisch omdat beleid wordt geïnitieerd, uitgevoerd en geëvalueerd op basis van cijfers verkregen met onderzoeksmethoden waarvan we tevoren kunnen betwijfelen of de resultaten voldoende valide zijn om goed beleid te waarborgen. Daarom zijn binnen de sociale wetenschappen diverse methoden van dataverzameling ontwikkeld waarmee onderrapportage zoveel mogelijk geminimaliseerd wordt. Eén van deze methoden is de randomized responsetechniek (RRT). Onderzoek wijst uit dat randomized response een zeer geschikte methode is om sensitief gedrag te meten. De crux van randomized response is dat de data gecontroleerd worden vervuild door het toevoegen van random error. Door deze random error wordt de privacy van de respondent volledig gewaarborgd, waardoor het minder lastig wordt om op gevoelige vragen eerlijk te antwoorden. Een voorbeeld:

U zit in een zaal met 100 mensen en er wordt u gevraagd uw hand op te steken als u ooit voor meer dan 25 euro materiaal van het werk mee naar huis heeft genomen voor privégebruik (papier, pennen, voedingsmiddelen)? Hoewel sommige mensen dit wel eens doen is het niet plezierig dit in het openbaar toe te geven, er zullen weinig handen omhoog gaan. De onderzoeker kan nu een RRT toepassen. Hij of zij vraagt de mensen in de zaal een munt te nemen en op te gooien. Als u kop heeft gegooid en /of ooit iets meegenomen hebt van uw werk, dan wordt u verzocht uw hand op te steken. Nu steekt

minstens de helft van de mensen in de zaal de hand op, dus personen die ooit iets meenamen van het werk voor privegebruik kunnen rustig ook de hand opsteken. Niemand weet immers waarom de hand wordt opgestoken, vanwege de munt of vanwege de vraag. Wanneer er 57 handen omhoog gaan, kunnen we vrij simpel uitrekenen dat ongeveer 14% van de aanwezigen wel eens iets heeft meegenomen van het werk.

Binnen randomized response bestaan een aantal varianten. In de *forced responsemethode* dient de respondent de vraag naar waarheid te beantwoorden, of hij wordt verplicht met ‘ja’ of ‘nee’ te antwoorden. Welke van deze drie antwoordmogelijkheden de respondent aangereikt krijgt wordt bepaald door een van tevoren vastgesteld kansmechanisme, bijvoorbeeld door het gooien van dobbelstenen. Deze methode – die uitgebreid in hoofdstuk 2 wordt besproken - is te combineren met een interviewtechniek die in hoofdstuk 3 wordt besproken, de computergestuurde zelfrapportage (CASI). De combinatie van de forced-responsemethode CASI en leidt tot de methode van computergestuurde randomized response zelfrapportage (CARR), die volgens de huidige inzichten wordt aangemerkt als de ‘current best’ methode voor het verzamelen van sensitieve data. In hoofdstuk 1 wordt aangegeven hoe deze methode in de praktijk kan worden toegepast.

Zoals uit het bovenstaande blijkt worden in deze handleiding een aantal methoden voor het verzamelen van sensitieve gegevens behandeld. Het eerste hoofdstuk behandelt de methode CARR. Aan de hand van een voorbeeld uit de praktijk wordt ingegaan op de vorm van de enquêtevragen en de instructies aan de respondenten. In het tweede hoofdstuk bespreken we de voor- en nadelen van een aantal varianten binnen randomized response. Hierbij ligt de nadruk vooral op de efficiëntie en de bepaling van de benodigde steekproefomvang bij deze varianten. In het derde hoofdstuk bespreken we een drietal alternatieven voor randomized response. Tenslotte geven we in het laatste hoofdstuk een overzicht van up-to-date literatuur die relevant is voor het verzamelen van sensitieve data.

1. Computergestuurde randomized response (CARR)

1.1 De 'current best' methode CARR

Meta-analyses¹ van onderzoeken waarbij de randomized responsemethode wordt vergeleken met standaard 'directe-vraagmethoden' wijzen uit dat het gebruik van de randomized responsetechniek meer valide schatters oplevert. Dit is speciaal het geval naarmate het onderwerp meer sociaal gevoelig of bedreigend is (Lensvelt-Mulders, Hox en Van der Heijden, opgestuurd). Dit maakt de randomized responsemethode zeer geschikt voor het doen van zelfrapportage-onderzoek naar regelovertrading. Omdat tegenwoordig de meeste zelfrapportages worden gedaan met behulp van computergestuurde vragenlijsten heeft de capaciteitsgroep 'Methodenleer en Statistiek' van de Universiteit van Utrecht in samenwerking met het NIPO, Amsterdam een methode ontwikkeld die beide vormen in zich verenigt, de computer gestuurde randomized response vragenlijst voor *zelfrapportage*, die we verder CARR (computer assisted randomized response) zullen noemen. Het doel van het ontwikkelen van een CARR-procedure is om te komen tot een computer gestuurde randomized response vragenlijst die de respondent geheel zelfstandig, zonder hulp van derden, van het begin tot het einde kan invullen.

Hieronder volgt een handleiding voor het opzetten van een CARR onderzoek, zoals uitgevoerd voor het onderzoek naar regelovertrading door uitkeringsgerechtigden uitgevoerd in 2002 door het NIPO in samenwerking met de Universiteit van Utrecht en in opdracht van het Ministerie van SoZaWe. De handleiding bestaat uit een beknopte beschrijving van de in het onderzoek gebruikte variant van de randomized-responsetechniek, de forced-responsmethode (voor uitgebreidere bespreking van deze methode wordt verwezen naar paragraaf 2.1.2). Daarna volgt de instructie voor de respondenten. In deze instructie zijn de resultaten uit twee cognitief kwalitatieve vooronderzoeken verwerkt (Boeije en Lensvelt-Mulders, 2002).

¹ Het statistisch vergelijken van onderzoeksresultaten van meerdere onafhankelijke studies.

De forced-responsemethode is gebruikt vanwege de combinatie van efficiëntie, uitleggemak en de goede resultaten die in eerder onderzoek zijn behaald met deze methode (van der Heijden et al., 1998). De forced-responsemethode werkt als volgt. De geïnterviewde wordt een vraag gesteld, waarbij slechts de antwoorden ‘ja’ en ‘nee’ mogelijk zijn. Voorafgaand aan de beantwoording gooit de geïnterviewde met twee dobbelstenen zonder dat iemand anders het resultaat van de worp te zien krijgt. De geïnterviewde moet op de gestelde vraag ‘geforceerd’ met *ja* antwoorden als de som van de worp 2, 3, of 4 is en ‘geforceerd’ met *nee* antwoorden als de som van de worp 11 of 12 is, de geïnterviewde moet ‘eerlijk’ antwoord geven wanneer de som van de worp tussen 5 en 10 ligt. Op deze manier weet niemand of de geïnterviewde de vraag ‘eerlijk’ dan wel ‘geforceerd’ heeft beantwoord.

1.2 Instructie vragenlijst

De volgende instructie is voorgelegd aan de respondenten in het NIPO /UU onderzoek voor het Ministerie van Sociale Zaken en Werkgelegenheid:

Wij willen u een paar vragen stellen over

Uit eerder onderzoek weten we dat veel mensen het moeilijk vinden dit soort vragen te beantwoorden, omdat ze dat te privé vinden. Sommige mensen zijn bang dat hun eerlijk antwoord negatieve gevolgen zal hebben. Maar wij willen niemand in verlegenheid brengen. Daarom hebben de Universiteit van Utrecht en het NIPO besloten om deze vragen te stellen met behulp van een omweg, die uw privacy volledig garandeerd.

U gaat straks de vragen beantwoorden met behulp van twee dobbelstenen. Daarmee kunt u 2 of 12 gooien of alles daartussenin. U antwoordt afhankelijk van het aantal ogen dat u gegooid hebt. Door deze omweg is uw privacy gegarandeerd! Niemand, de enquêteur niet, de onderzoekers niet en ook niet de uitkeringsinstantie kan ooit te weten komen wat u precies heeft geantwoord.

Hoe werkt dat dan?

^AU ziet op het scherm twee rollende dobbelstenen verschijnen. Door op ‘enter’ te drukken kunt u de dobbelstenen stoppen. U ziet dan direct het aantal ogen dat u gegooid heeft. Als u daarna nog een keer op ‘enter’ drukt komt de vraag op uw scherm.

Als u 2, 3, of 4 hebt gegooid drukt u altijd op 1 (= ja)

Als u 11 of 12 hebt gegooid drukt u altijd op twee (=nee)

Als u 5, 6, 7, 8, 9, of 10 hebt gegooid dan beantwoordt u de vraag altijd naar waarheid. U antwoordt ‘ja’ door op 1 te drukken of ‘nee’ door op 2 te drukken.

U heeft twee dobbelstenen gekregen. Daarmee gooit u een keer en u telt het aantal ogen dat u totaal gegooit heeft.

Als u 2, 3, of 4 hebt gegooit drukt u altijd op 1 (= ja)

Als u 11 of 12 hebt gegooit drukt u altijd op twee (=nee)

Als u 5, 6, 7, 8, 9, of 10 hebt gegooit dan beantwoordt u de vraag altijd naar waarheid. U antwoordt 'ja' door op 1 te drukken of 'nee' door op 2 te drukken.

Ook al vindt u het misschien een beetje een rare techniek, de dobbelsteenmethode is leuk om te gebruiken en nuttig omdat ze ervoor zorgt dat uw privacy gegarandeerd is. Omdat niemand anders dan uzelf weet wat u gegooit heeft, weet ook niemand waarom u 1 of 2 hebt gedrukt. Uw ware antwoord is dus echt geheim. Toch is het nuttig omdat de onderzoekers van de Universiteit van Utrecht een schatting kunnen maken van het aantal mensen in de hele groep die '1' hebben geantwoord vanwege het aantal ogen dat ze hebben gegooit, en het aantal mensen dat '1' heeft geantwoord omdat ze eerlijk antwoord hebben gegeven.

Nu volgen vier oefenvragen:

1 Gooit u alstublieft eerst de dobbelstenen (*bij gebruik van virtuele dobbelstenen komen deze automatisch als eerste in beeld*)

Als u 2, 3, of 4 hebt gegooit druk dan op 1 (= ja)

Als u 11 of 12 hebt gegooit druk dan op 2 (=nee)

Als u 5, 6, 7, 8, 9, of 10 hebt gegooit beantwoord dan de volgende vraag:

Hebt u vandaag de krant gelezen?

1 = ja

2 = nee

2 Gooit u alstublieft eerst de dobbelstenen (*bij gebruik van virtuele dobbelstenen komen deze automatisch als eerste in beeld*)

Als u 2, 3, of 4 hebt gegooit druk dan op 1 (= ja)

Als u 11 of 12 hebt gegooit druk dan op 2 (=nee)

Als u 5, 6, 7, 8, 9, of 10 hebt gegooit beantwoord dan de volgende vraag:

Heeft u de afgelopen 12 maanden wel eens zwart (dus zonder geldig kaartje) gereisd met de bus, tram of trein?

1 = ja

2 = nee

3 Gooit u alstublieft eerst de dobbelstenen (*bij gebruik van virtuele dobbelstenen komen deze automatisch als eerste in beeld*)

Als u 2, 3, of 4 hebt gegooid druk dan op 1 (= ja)

Als u 11 of 12 hebt gegooid druk dan op 2 (=nee)

Als u 5, 6, 7, 8 ,9, of 10 hebt gegooid beantwoord dan de volgende vraag:

Hebt u de afgelopen 12 maanden wel eens door rood licht gereden?

1 = ja

2 = nee

^B4 Gooit u alstublieft eerst de dobbelstenen (*bij gebruik van virtuele dobbelstenen komen deze automatisch als eerste in beeld*)

Als u 2, 3, of 4 hebt gegooid druk dan op 1 (= ja)

Als u 11 of 12 hebt gegooid druk dan op 2 (=nee)

Als u 5, 6, 7, 8 ,9, of 10 hebt gegooid beantwoord dan de volgende vraag:

Is of was uw moeder jarig in januari?

1 = ja

2 = nee

^CU heeft net alle oefenvragen gemaakt. Misschien is het u overkomen dat u 2, 3, of 4, gooide en daarom 1 (=ja) moest indrukken, terwijl uw echte antwoord ‘nee’ zou zijn geweest. Iemand die bijvoorbeeld nooit door het rode licht rijdt, moest nu misschien toch ‘ja’ invullen. Of dat u 11 of 12 gooide en dus 2 (=nee) moest invullen terwijl uw echte antwoord ‘ja’ zou zijn geweest. Iemand die bijvoorbeeld vanochtend wel de krant heeft gelezen, moest misschien toch ‘nee’ antwoorden. Uit eerder onderzoek weten we dat sommige mensen dat vreemd vinden en soms vinden dat ze een onjuist en zelfs oneerlijk antwoord geven. Hierover hoeft u zich geen zorgen te maken. Bij deze dobbelsteenmethode bent u juist eerlijk als u antwoord geeft volgens de regels. Het is net als bij een spel, als u een spel speelt volgens de regels dan speelt u het spel eerlijk.

^DDan volgen nu onze vragen

1.3 Toelichting op de instructies

Sommige delen van de instructietekst zijn voorzien van een letter in superschrift, de betekenis daarvan willen we hier even kort toelichten.

^A In voornoemd NIPO /UU onderzoek werd gebruik gemaakt van virtuele dobbelstenen, die waren opgenomen in het enquêteprogramma. De dobbelstenen werden aangeboden in een window, aan het begin van elke nieuwe vraag. Wanneer het programma begon 'rolde' de dobbelstenen met grote snelheid. Door op 'enter' te drukken, werden de dobbelstenen gestopt en verscheen het aantal gegooide ogen in beeld. Door nogmaals op enter te drukken werd de vraag in beeld gebracht. Vooronderzoek wees uit dat de proefpersonen groot vertrouwen hadden in de virtuele dobbelstenen, dat zij ze makkelijker vonden om te gebruiken dan de echte dobbelstenen en dat zij niet probeerde om af te drukken op een 'goede' worp, zoals dat in spelletjes op TV wel gebeurt. Dit laatste was overigens niet mogelijk omdat de snelheid waarmee de stenen rolden groter was dan het menselijk reactievermogen (Lensvelt-Mulders en Boeije, 2002).

^B Deze vraag kan worden gebruikt om te controleren of de respondenten de instructie goed hebben begrepen en ook hebben opgevolgd. In Nederland is ongeveer 1 op de 12 van de mensen jarig in Januari, hetgeen overeenkomt met 8.3 % van de totale bevolking. Juiste toepassing van de schattingsregels zou dus moeten resulteren in een schatting van .83 'ware' ja-antwoorden.

^C Deze instructie is later toegevoegd naar aanleiding van de resultaten van het kwalitatief vooronderzoek, waaruit bleek dat de proefpersonen het geforceerd antwoorden volgens de dobbelsteenregels vooral als oneerlijk ervoeren. Dit gold iets sterker voor het gedwongen ja-antwoorden dan het gedwongen nee-antwoord. Voor sommige proefpersonen was dit zelfs aanleiding om de regels niet meer te volgen (vals spelen). Het toevoegen van deze extra geruststelling na het maken van de oefenvragen maakte dat in de tweede ronde van het vooronderzoek alle proefpersonen de regels van de RRT wel netjes volgde.

^D Het is aan te raden een aantal vragen zo te formuleren dat het nee-antwoord het gevoelige antwoord is. Hierdoor wordt voorkomen dat de respondenten na een aantal vragen ingevuld te hebben een zekere tegenzin ontwikkelen om 1 = 'ja' te drukken. Natuurlijk moet daarbij voorkomen worden dat een vraag dubbel negatief wordt gesteld. Een voorbeeld van een vraag waarbij 'nee' het bezwarende antwoord is zou bijvoorbeeld kunnen zijn: *'Heeft u altijd volledige informatie gegeven over uw financiële zaken?'*

Dit wisselen van perspectief maakt het noodzakelijk extra aandacht te geven aan de dataverwerking, omdat antwoorden van randomized response onderzoeken niet eenvoudigweg kunnen worden omgecodeerd. Immers het aantal geforceerde nee-antwoorden is niet gelijk aan het aantal geforceerde ja-antwoorden, waardoor de berekening van de steekproefgemiddelden en hun varianties moet worden aangepast.

2. Alternatieve technieken binnen randomized response

In dit hoofdstuk bespreken we naast de reeds genoemde forced-responsetechniek ook enkele alternatieve randomized responsetechnieken (RRT). Voor deze methoden geven we aan hoe op basis van de gegevens uit de steekproef een schatting van de prevalentie van het sensitieve gedrag in de populatie gemaakt kan worden en hoe de variantie deze schatting berekend kan worden. Vervolgens laten we zien hoe de variantie samenhangt met betrouwbaarheid van de schatting en de steekproefomvang. Tenslotte geven we aan wat ons inziens een bevredigende betrouwbaarheid is en hoe groot de steekproef moet zijn om een dergelijke betrouwbaarheid te realiseren.

2.1 Randomized Responsetechnieken

In 1965 publiceerde Warner's voor het eerst over de randomized responsetechniek en sindsdien zijn er meer dan 300 artikelen over deze dataverzamelmethode verschenen. Daarin worden o.a. aanpassingen in Warner's methode beschreven die de methode statistisch efficiënter maken, zodat minder grote steekproeven nodig zijn om tot betrouwbare resultaten te komen. Of de randomized responsetechniek naast statistisch efficiënt ook zuiver is, hangt af van de validiteit van de resultaten, en die hangt volledig af van de bereidheid van de respondenten om de procedure daadwerkelijk te volgen.

In het onderstaande willen we vier succesvolle randomized responsemethoden beschrijven. Uit literatuurstudie blijkt dat in 92% van al het randomized response onderzoek gebruik wordt gemaakt van slechts 4 verschillende technieken, te weten de *methode van Warner (1965)*, de forced-responsemethode, het *unrelated question design (UQD)* en *Kuk's kaartmethode*. De forced-responsemethode en Kuk's kaartmethode heeft de onderzoeksgroep van de Universiteit van Utrecht met succes toegepast in onderzoek naar uitkeringsfraude. Van al de bovengenoemde vormen van de randomized responsemethode zullen we naast een korte beschrijving ook de vergelijkingen geven voor het berekenen van populatiegemiddelden en hun varianties, een beschrijving van de sterke en zwakkere kanten van de

methode en een overzicht van de literatuur waarbij deze methoden worden gebruikt. In het navolgende gaan we bij het berekenen van populatiegemiddelden en varianties er steeds van uit dat de steekproef een a-selecte trekking is uit de populatie.

2.1.1 Warner's originele methode

In 1965 ontwikkelde Warner een onderzoekstechniek die, door het totaal waarborgen van de privacy van respondenten, zou moeten leiden tot een vermindering van non-response en van het geven van sociaal wenselijke antwoorden op gevoelige vragen. Een respondent krijgt twee statements voorgelegd, bijvoorbeeld:

- Ik ben voor de doodstraf
- Ik ben tegen de doodstraf

Met behulp van een kansspelletje (met bijvoorbeeld dobbelstenen, speelkaarten of muntjes) wordt nu bepaald welk statement moet worden beantwoord met 'waar' of 'niet waar'. Omdat alleen de respondent weet welk statement hij of zij heeft beantwoord, heeft het individuele antwoord ook nu geen betekenis. 'Waar' kan immers op beide statements betrekking hebben. Echter het is statistisch mogelijk om de prevalentie van het gevoelige gedrag in de populatie te schatten met behulp van de steekproefschatter $\hat{\pi}$:

$$\hat{\pi} = \frac{\lambda + (p-1)}{2p-1} \quad (1)$$

waar λ de geobserveerde proportie 'waar'-antwoorden in onze steekproef is, p de kans dat de respondent het eerste statement moet beantwoorden (p is niet gelijk aan .5), en $p-1$ de kans dat het tweede statement moet worden beantwoord. De variantie is gelijk aan

$$\text{var}(\hat{\pi}) = \frac{\hat{\pi}(1-\hat{\pi})}{n} + \frac{p(1-p)}{n(2p-1)^2} \quad (2)$$

Deze variantie neemt toe als het aantal individuen (n) in de steekproef klein is en als p dichter bij 0,5 ligt. Vergelijking 2 laat zien dat de RRT van Warner veel minder efficiënt is dan gewone directe-vraag designs, waarvan de term $\pi(1-\pi)/n$ in vergelijking 2 de variantie is. Aan deze variantie wordt als gevolg van het gebruik van randomized response de variantie $\frac{p(1-p)}{n(2p-1)^2}$ toegevoegd, waar uit blijkt dat

in Warner's methode grotere steekproeven nodig zijn dan in het gewone directe-vraag design om dezelfde

betrouwbaarheidsintervallen rondom het gemiddeld te krijgen. De berekening van de benodigde omvang van de steekproef in de verschillende methoden wordt behandeld in paragraaf 2.3. Warner's methode is gebruikt in onderzoek naar academische fraude (Schotland, 1982).

2.1.2 De forced-responsemethode

Zoals eerder gezegd, is het bij alle randomized response designs belangrijk om de data op een gecontroleerde manier te 'vervuilen' door het toevoegen van een vastgestelde proportie random error. Bij het gebruik van de forced-responsemethode wordt dit gedaan door de respondent in een bekend aantal gevallen te forceren tot het geven van een 'ja'- of 'nee'-antwoord. Alle respondenten krijgen daartoe dezelfde vragen voorgelegd. Met behulp van een randomizer (bijvoorbeeld 2 dobbelstenen) wordt bepaald welk antwoord de respondent op die vraag gaat geven, geforceerd 'ja', geforceerd 'nee', of de vraag naar waarheid beantwoorden.

Als een respondent bijvoorbeeld 2, 3 of 4 gooit dan moet hij of zij 'ja' antwoorden, ongeacht het ware antwoord op de vraag. Als bijvoorbeeld de vraag is: *Heeft u de laatste 7 dagen een of meerdere keren harder dan 120 km/uur gereden op de snelweg?*, en de respondent gooit 2, 3, of 4 dan is het een antwoord een geforceerd 'ja', ook als de respondent nooit harder dan 120 km/uur heeft gereden, zelfs als hij of zij niet eens een rijbewijs heeft. Als de respondent 11 of 12 gooit dan is het antwoord een geforceerd 'nee', ook als de respondent wel degelijk een of meerdere keren harder dan 120 km/uur heeft gereden. Als de respondent 5, 6, 7, 8, 9 of 10 gooit moet de vraag naar waarheid beantwoord worden, dat wil zeggen als de respondent harder dan 120 km/uur gereden heeft met 'ja' en als hij of zij niet harder dan 120 km/uur heeft gereden met 'nee'.

Met twee dobbelstenen is kans op 4 of minder ogen 1/6 en op 11 of 12 ogen 1/12. Dit betekent dat gemiddeld $1 - (1/6 + 1/12) = 3/4$ van de respondenten de vraag naar waarheid moet beantwoorden. De overige 25% van de data bestaat uit verplichte antwoorden en is dus 'vervuild'. Aangezien echter precies bekend is hoe de vervuiling tot stand is gekomen, kan ervoor worden gecorrigeerd. Daarom is het mogelijk aan de hand van de zuivere steekproefschatter (Fox en Tracy, 1986)

$$\hat{\pi} = \frac{\hat{\lambda} - \theta}{p} \quad (3)$$

een schatting te maken van de prevalentie van de overtreding in de populatie met als variantie:

$$\text{var}(\hat{\pi}) = \frac{\hat{\lambda}(1 - \hat{\lambda})}{np^2} \quad (4)$$

waarbij $\hat{\pi}$ staat voor de geschatte percentage hardrijders gedurende de laatste 7 dagen is, λ de proportie gegeven ‘ja’ antwoorden, θ de proportie geforceerde ‘ja’ antwoorden en p de kans op een antwoord naar waarheid. Het grote voordeel van de forced-responsemethode is de grotere efficiëntie in vergelijking met andere randomized-responsetechnieken. Dat wil zeggen dat de steekproef minder groot hoeft te zijn dan bij Warner’s methode.

Een nadeel van deze methode is dat de respondent wordt geforceerd tot oneerlijkheid. Omdat het ‘ja’ of ‘nee’ antwoord gekoppeld is aan een vraag lijkt het er voor de respondent op dat hij of zij toegeeft het gevoelige gedrag te hebben vertoond. Daardoor kan de neiging ontstaan om de regels van de randomized responsemethode niet meer te volgen (Boeije, en Lensvelt-Mulders, 2002). In ons eigen onderzoek hebben we dit opgelost door het invoegen van een extra instructie nadat de respondent de oefenvragen heeft gemaakt en door steeds eerst de instructie te geven en pas daarna de vraag te stellen. Hierdoor wordt de randomized response regel benadrukt en psychologisch meer losgekoppeld van de gevoelige vraag.

Als u 2, 3, of 4 hebt gegooid druk dan op 1 (= ja)

Als u 11 of 12 hebt gegooid druk dan op 2 (=nee)

Als u 5, 6, 7, 8 ,9 of 10 hebt gegooid beantwoord dan de volgende vraag:

Hebt u vandaag de krant gelezen?

1 = ja

2 = nee

De forced-responsemethode is recentelijk met goede resultaten gebruikt in onderzoek naar fraude in organisaties (Armacost et al., 1991) onderzoek naar regelovertreding (Van der Heijden et al., 1998, 2000; Elffers, Van der Heijden, en Hezewijk, opgestuurd), diefstal in organisaties (Wimbush et al. 1997) en in naar regelovertreding in het kader van de Warenwet, de Meststoffenwetgeving en de Wet Verontreiniging Oppervlaktewateren.

2.1.3 Unrelated Question design (UQD)

Dit design lijkt sterk op het forced-responsemethode, met dien verstande dat nu niet een geforceerd antwoord moet worden gegeven, maar dat er een neutrale vraag wordt beantwoord als de som van een worp 2, 3, 4, 11 of 12 is en de gevoelige vraag als de som van de worp 5, 6, 7, 8, 9 of 10 is.

Gooit u eerst met de dobbelstenen.

Als u 2, 3, 4, 11 of 12 hebt gegooid, beantwoord dan de volgende vraag:

Leest u elke dag de krant?(of: is uw moeder jarig tussen 1 januari en 1 april?)

1=ja

2=nee

Als u 5, 6, 7, 8, 9 of 10 hebt gegooid beantwoord dan de volgende vraag:

Heeft u wel eens een herstel van uw klachten..... bemerkt zonder dat u dit aan de uitkeringsinstantie heeft doorgegeven?

1=ja

2=nee

Het *unrelated question design* (UQD) kent twee vormen, één waarbij het voorkomen van de neutrale vraag in de populatie niet bekend is (bijvoorbeeld ‘*Leest u elke dag de krant?*’), en een waarbij dit wel tevoren bekend is (bijvoorbeeld ‘*is uw moeder jarig tussen 1 januari en 1 april?*’). Deze laatste vorm is het meest toegepast en is statistisch vergelijkbaar met de forced-responsemethode, immers de onderzoeker weet tevoren hoeveel respondenten verplicht ‘ja’ of ‘nee’ gaan antwoorden (25 % van de respondenten beantwoordt de verjaardagsvraag en daarvan zal 25% bevestigend moeten antwoorden, omdat in drie maanden 25% van de bevolking en dus ook van de moeders jarig is). Wanneer we de kans op een ‘ja’-antwoord op de ongerelateerde vraag met π_j aanduiden dan wordt de prevalentie van het sensitieve gedrag op vergelijkbare wijze geschat als bij forced response, waarbij de θ vervangen wordt door $(1-p)\pi_j$,

$$\hat{\pi} = \frac{\lambda - (1-p)\pi_j}{p}.$$

De variantieschatting is identiek aan die van forced response.

Deze vorm van het UQD heeft twee voordelen. Ten eerste is het samen met de forced-response methode de meest efficiënte vorm van de randomized-responsetechnieken (kleinere steekproeven nodig voor gelijke betrouwbaarheidsintervallen als vraag-antwoord designs). Ten tweede is deze vorm van randomized response voor de respondenten makkelijk te begrijpen, waardoor het vertrouwen in de methode toeneemt, en daardoor de validiteit van de resultaten. Een groot nadeel is echter dat voor elke gevoelige vraag een nieuwe neutrale vraag met bekende harde populatie gegevens moet worden gemaakt, in de praktijk blijkt dit zeer lastig te zijn.

De andere vorm van het UQD is die waarbij een willekeurige neutrale vraag wordt gesteld en het neutrale gedrag dus ook geschat moet worden. In dat geval moet de steekproef in twee subgroepen worden opgesplitst. Subgroep 1 beantwoordt de gevoelige vraag met kans p en de neutrale vraag met kans $1-p$. Subgroep 2 daarentegen beantwoordt de neutrale vraag met kans p en de gevoelige vraag met kans $1-p$. Op die manier kan een schatter van beide populaties worden berekend, samen met de bijbehorende varianties.

Het voorkomen in de populatie van het gevoelige gedrag kan worden geschat als:

$$\hat{\pi}_a = \left[\hat{\lambda}_1(1-p_2) - \hat{\lambda}_2(1-p_1) \right] / (p_1 - p_2) \quad (5)$$

waar λ_1 de geobserveerde proportie 'ja'-antwoorden is in het eerste subgroep en λ_2 is de geobserveerde proportie 'ja'-antwoorden in het tweede subgroep, met als variantie

$$\hat{\text{var}}(\hat{\pi}) = \frac{1}{(p_1 - p_2)^2} \left[\frac{\hat{\lambda}_1(1-\hat{\lambda}_1)(1-p_2)^2}{n_1} + \frac{\hat{\lambda}_2(1-\hat{\lambda}_2)(1-p_1)^2}{n_2} \right] \quad (6)$$

Het voordeel van dit design is dat de respondenten het principe van dit design heel goed kunnen begrijpen, en dat het aantal mogelijke neutrale vragen zeer groot is. Het grote nadeel van dit design is dat het erg inefficiënt is, er zijn zeer grote steekproeven nodig om een acceptabel betrouwbaarheidsinterval te verkrijgen.

Het UQD is onder andere succesvol gebruikt voor het schatten van academische fraude (Scheers, 1987) en het schatten van de omvang van de populatie drugsgebruikers onder tieners (Brewer, 1981). Ook in Nederland is het UQD toegepast bij onderzoek naar de Wet Individuele Huursubsidie (IHS) en de

Bestrijdingsmiddelenwet (BMW) (Elffers, Van der Heijden en Hezemans), helaas niet met goed resultaat. Dit laatste was waarschijnlijk het gevolg van een ongelukkige toepassing, waarbij elke gevoelige vraag gekoppeld was aan dezelfde neutrale vraag. Hierdoor werden de respondenten achterdochtig dat de onderzoeker door een bepaald patroon in de vragen de juiste antwoorden kon anticiperen.

2.1.4 Kuk's kaartmethode

Bij veel randomized response methodes wordt van de respondent verwacht dat hij of zij een ja-antwoord geeft bij een gevoelige vraag. Dit ja-antwoord is altijd bedreigend, en dit kan de respondent hinderen bij het geven van een eerlijk antwoord (Fox en Tracy, 1986). Bij Warner's methode wordt dit voorkomen doordat de statements zo geformuleerd zijn dat men met waar of niet-waar moet antwoorden of dat zowel 'ja' als 'nee' antwoorden bedreigend kunnen zijn. Kuk heeft in 1990 een methode ontwikkeld waarbij het niet langer noodzakelijk is om 'ja/nee' of 'waar/niet waar' te antwoorden. Kuk's design is statistisch gelijk aan Warner's oorspronkelijke randomized response techniek, maar psychologisch sterker. De respondent heeft twee stapels speelkaarten voor zich liggen, een *ja* en een *nee* stapel. De ja-stapel bevat 80% rode kaarten en de nee-stapel bevat 20% rode kaarten. Voordat de vraag gesteld wordt, trekt de respondent uit elke stapel een kleur en hij noemt de kleur die overeenkomt met de stapel van het juiste antwoord.

De proportie rode kaarten is θ_1 in de ja-stapel en θ_2 in de nee-stapel, terwijl θ_1 niet gelijk is aan θ_2 . Het antwoord dat gegeven wordt is dus de kleur van de getrokken kaart uit de juiste stapel. De schatter van π wordt gegeven door

$$\hat{\pi} = (r - \theta_1) / (\theta_1 + \theta_2) \quad (7)$$

waarbij r de proportie geobserveerde rode kaarten uit een random steekproef van n respondenten is. De variantie van $\hat{\pi}$ wordt gegeven door

$$\text{var}(\hat{\pi}) = \frac{\hat{\phi}(1 - \hat{\phi})}{n(\theta_1 - \theta_2)^2} \quad (8)$$

waarbij

$$\hat{\phi} = \hat{\pi}\theta_1 + (1 - \hat{\pi})\theta_2 \quad (9)$$

Deze methode is psychologisch erg sterk. Individuen hoeven zichzelf niet te beschuldigen door ja te antwoorden. In een vergelijking tussen Kuk's kaartmethode, de forced-responsemethode en CAI kwam Kuk als meest valide methode naar voren, het verschil tussen de forced-responsemethode en Kuk's kaartmethode was echter niet significant (van der Heijden et al.1998, 2000). Het grote nadeel van het gebruik van de methode van Kuk is de inefficiëntie (grote standaardfouten) waardoor veel grotere steekproeven nodig zijn.

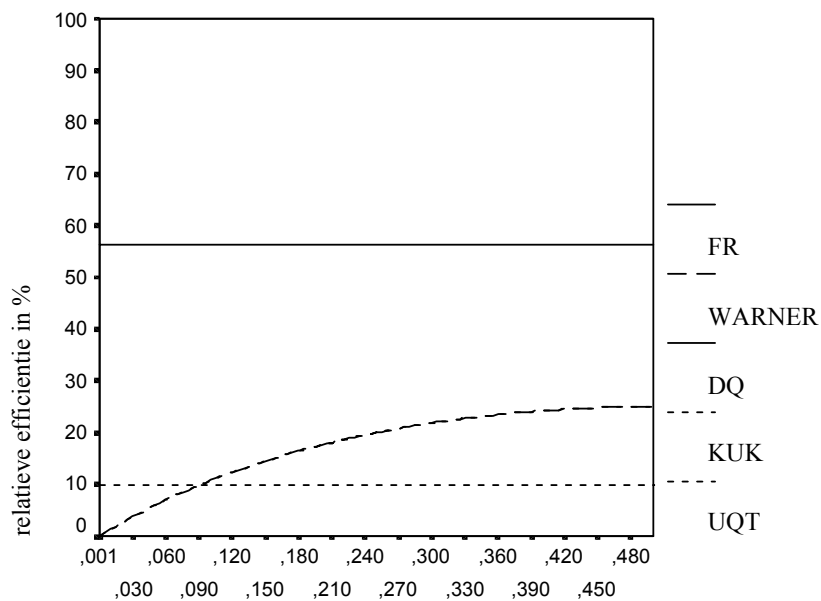
2.2 Relatieve efficiëntie

Bij een randomized response steekproef speelt het toeval op twee manieren een rol. In de eerste plaats hadden andere mensen in de steekproef terecht kunnen komen dan de mensen die er nu inzitten. In de tweede plaats hadden de dobbelstenen anders kunnen rollen. Het zal duidelijk zijn dat de toevalsfluctuatie in een randomized respons onderzoek daarom groter is dan bij het gebruik van de directe-vraagmethode, hetgeen zich uit in een grotere variantie van de schatting van de randomized responsemethoden. In randomized response kan de variantie van een schatting in principe op twee manieren geminimaliseerd worden, enerzijds door de steekproef zo groot mogelijk te nemen en anderzijds door de parameters in de randomized responsemethode zo optimaal mogelijk te kiezen (dit laatste uiteraard zonder dat de validiteit van de methode verloren gaat). Door de kans op de gevoelige vraag 75 % te maken ($p = .75$ geeft voldoende objective én subjectieve bescherming aan de respondenten) en de geforceerde ja-antwoorden (in forced response of met behulp van een neutrale vraag) in dezelfde orde van grootte te kiezen als het verwachte gevoelige gedrag wordt de variantie geminimaliseerd.

De varianties van methoden kunnen onderling vergeleken worden door de relatieve efficiëntie van een methode t.o.v. een referentiemethode te berekenen. De efficiëntie van een randomized responsetechniek (RRT) t.o.v. de directe-vraagmethode (DQ) wordt gegeven door $\frac{\text{var}(DQ)}{\text{var}(RRT)} \times 100\%$. In onderstaande

grafiek zijn de relatieve efficiënties van vijf methoden t.o.v. van de directe-vraagmethode uitgezet tegen de prevalentie van het gevoelig gedrag, waarbij deze laatste varieert tussen .001 (1 % van de populatie heeft de regel overtreden) en .5 (50% van de populatie heeft de regel overtreden).

Figuur: Relatieve efficiëntie randomized-responsetechnieken



voorkomen in populatie wordt variabel tussen .001 and .5

Het meest efficiënte design is de directe-vraagmethode zelf met een efficiëntie van 100%. Van de randomized responsemethoden is de forced-responsemethode (en dus ook het statistisch vergelijkbare UQD met een neutrale vraag waarvan de prevalentie in de populatie *bekend* is) het meest efficiënte design, de met een relatieve efficiëntie van 58 % voor elke willekeurige prevalentie in de populatie. De efficiëntie van Kuk's methode en de efficiëntie van Warners' methode zijn aan elkaar gelijk (de kromme lijn, beide lijnen vallen samen), en deze nemen toe, naarmate ook de prevalentie toeneemt. De maximale haalbare efficiëntie van deze methoden wordt bereikt bij een prevalentie van 50% en bedraagt circa 26%. Bij prevalenties onder de 8% is de efficiëntie van beide methoden lager dan de 10% van het UQD met neutrale vraag waarvan het voorkomen in de populatie onbekend is.

2.3 Betrouwbaarheidsinterval en steekproefgrootte

In de vorige paragraaf hebben we met behulp van de varianties laten zien hoe de steekproefomvang van diverse methodes ten opzichte van elkaar bepaald kan worden om een gelijke betrouwbaarheid te krijgen. Deze procedure geeft echter geen antwoord op de vraag hoe groot de steekproefomvang moet zijn om een bevredigende betrouwbaarheid te krijgen. In deze paragraaf laten we zien hoe aan de hand van een 95% betrouwbaarheidsinterval en de tabellen in bijlage A de minimaal benodigde steekproefomvang kan worden bepaald. Het 95%-betrouwbaarheidsinterval wordt gegeven door $\hat{\pi} \pm 1.96\sqrt{\widehat{\text{var}}(\hat{\pi})}$. De beide waarden die met deze vergelijking worden verkregen zijn de onder- en bovengrens waarbinnen het populatiegemiddelde zich met 95% zekerheid bevindt. Dit betekent dat de kans dat de ware prevalentie in de populatie meer dan $\pm 1.96\sqrt{\widehat{\text{var}}(\hat{\pi})}$ afwijkt van de geschatte prevalentie in de steekproef slechts 5% bedraagt.

In bijlage A zijn voor de forced-response methode, Kuk en de directe-vraagmethode tabellen opgenomen waarmee op eenvoudige wijze het betrouwbaarheidsinterval kan worden berekend bij een bepaalde prevalentie π en een bepaalde steekproefomvang n . De waarden in de cellen van de tabel zijn de getallen gegeven door $1.96\sqrt{\widehat{\text{var}}(\hat{\pi})}$, door deze getallen bij de waarde van π op te tellen en af te trekken wordt het 95% betrouwbaarheidsinterval gevonden. Zo is bijvoorbeeld af te lezen dat bij een steekproefomvang van 400 en een steekproefgemiddelde van 20% het betrouwbaarheidsinterval van 13.02% tot 26.08% loopt (20% +/- 6.08%).

Uit de tabellen is af te leiden dat de grootte van het betrouwbaarheidsinterval niet alleen afhangt van de steekproefomvang maar ook van het steekproefgemiddelde, het interval is het grootst bij waarden rond de 50% en wordt kleiner naarmate het steekproefgemiddelde meer van 50% afwijkt. Echter, een interval van 10% rondom een waarde van $\hat{\pi} = .40$, dus $\{.35, .45\}$, is relatief gezien nauwkeuriger dan het interval van 10% $\{.05, .15\}$ rondom een $\hat{\pi} = .10$. Daardoor kan bij de prevalenties rond de 50% met kleinere steekproeven kan worden volstaan om een bevredigend betrouwbaarheidsinterval te krijgen, en zijn grotere steekproeven nodig naarmate de prevalentie meer van de 50% afwijkt.

Uit het voorgaande blijkt dat bij het vaststellen van een minimale steekproefomvang een redelijke inschatting van de prevalentie van het sensitieve gedrag in de populatie gewenst is. Nu zal dat in de praktijk vaak een probleem zijn, het onderzoek wordt immers juist gedaan omdat dat de prevalentie van het sensitieve gedrag onbekend is. Een mogelijke oplossing is om bij het berekenen van de minimale van steekproefomvang uit te gaan van een prevalentie van 7% en een betrouwbaarheidsinterval van +/- 4%,

dus van 3% tot 11%. Door de prevalentie op voorhand relatief laag in te schatten en de ondergrens van het betrouwbaarheidsinterval op enige afstand van 0% wordt voorkomen dat een te kleine steekproef wordt samengesteld en dat de resultaten achteraf onbruikbaar blijken te zijn. In tabel 1 van bijlage A is te zien dat voor de forced-responsemethode dit criterium bij een steekproefomvang van ongeveer 750 personen wordt bereikt. In de overige twee tabellen is te zien dat om hetzelfde resultaat te halen bij de Kuk-methode circa 1250 personen benodigd zijn en bij de directe-vraagmethode slechts 150 personen.

2.4 Steekproefgrootte in eindige populaties

Bij de bepaling van de steekproefomvang in de vorige paragraaf werd steeds impliciet uitgegaan van een oneindig grote populatie met een grote, onbekende omvang. In sommige gevallen kan er echter sprake zijn van een eindige populatie met een bekende en geringe omvang, zoals bijvoorbeeld personen die gebruik maken van een bijzondere regeling of alle bedrijven binnen een bepaalde bedrijfstak. Voor eindige populaties is, in vergelijking met oneindige populaties, in principe een kleinere steekproefomvang benodigd om tot dezelfde betrouwbaarheidsintervallen te komen. De reden hiervoor is dat de variantie als gevolg van toevalsfluctuaties in de steekproef afneemt naarmate een groter percentage van de populatie in de steekproef is opgenomen. We kunnen het betrouwbaarheidsinterval van een eindige populatie weer berekenen uit de variantie. Voor een eindige populatie van omvang N en een steekproef van omvang n is in de forced-responsemethode de variantie van het steekproefgemiddelde gelijk aan

$$\text{var}(\hat{\pi}) = \text{var}(RR) + \left[1 - \frac{n-1}{N-1}\right] \frac{\hat{\pi}(1-\hat{\pi})}{n} \quad (10)$$

waarbij $\frac{1}{n} \left[\frac{\lambda - 2\theta\lambda + \theta^2}{p^2} - \frac{\lambda - \theta}{p} \right]$ de variantie is als gevolg van het forced-responsemethode en is

$\frac{\hat{\pi}(1-\hat{\pi})}{n}$ de steekproefvariantie (deze laatste is gelijk aan de variantie van de directe-vraagmethode). De

term $1 - \frac{n-1}{N-1}$ is de correctie voor het feit dat de populatie eindig is.

In de tabel van bijlage B is met behulp van vergelijking (10) voor de waarden $\hat{\pi} = 0.05, 0.1, 0.2$ en 0.4 de waarde $1.96\sqrt{\text{var}(\hat{\pi})}$ berekend bij verschillende combinaties van populatieomvang N en steekproefomvang n . Met deze tabel kan voor een eindige populatie op analoge wijze als in de vorige paragraaf de minimaal benodigde steekproefomvang bepaald worden. Uit de tabel blijkt dat het alleen zinnig is om randomized responseonderzoek op zeer kleine populaties (van circa 400 of kleiner) uit te

voeren indien de prevalentie van het sensitieve gedrag relatief hoog wordt ingeschat. Maar zelfs dan zal het vaak noodzakelijk zijn om de volledige populatie in de steekproef op te nemen. Verder is te zien dat bij de wat grotere populaties vaak kan worden volstaan met een iets kleinere steekproef, en dit effect is het grootst wanneer de prevalentie in de populatie de 50% nadert.

3. Alternatieven voor randomized responsetechnieken

3.1 Het computer gestuurde interview (CASI)

Op dit moment is de meest courante onderzoeksmethode om gevoelige onderwerpen te bevragen het gebruik van computer gestuurde zelfrapportage. Uit vergelijkend onderzoek blijkt dat een volledig computergestuurd zelf-interview meer openheid in de antwoorden oplevert dan een door de respondent zelf ingevulde papieren vragenlijst (De Leeuw, Hox, & Snijkers, 1995; Weisband & Kiesler, 1996; De Leeuw, 2001). Hierbij is het van belang dat *in de optiek van de respondent* de antwoorden volstrekt vertrouwelijk zijn. Zo vond Beebe (1998) dat in een onderzoek naar drugsgebruik via CASI (Computer Assisted Self Interview) de afstand tussen computers in een computerruimte van cruciaal belang was voor de openheid in antwoord. Wanneer computers te dicht bij elkaar staan, of wanneer de mogelijkheid bestaat dat anderen meekijken, wordt het gevoel van privacy van de respondent aangetast.

CASI levert dan ook meer openheid op dan computer gestuurde interviews waarbij de interviewer persoonlijk aanwezig is. Dit probleem probeert de interviewer te ondervangen door bij gevoelige vragen de computer aan de respondent te geven, die dan in alle privacy zelf deze vragen beantwoordt en daarna de PC teruggeeft aan de interviewer.

Functioneel analfabetisme kan de bruikbaarheid van CASI voor specifieke doelgroepen beperken. Ook bij lager opgeleiden en/of respondenten waarbij het Nederlands niet de moedertaal is kunnen leesproblemen de meetfout vergroten en zo de validiteit van de antwoorden aantasten. Speciaal voor deze groepen is A-CASI ontworpen: Audio-Computer Assisted Self Interview. In deze toepassing hoort de respondent via een koptelefoon de vraag en mogelijke antwoordcategorieën op het moment dat deze ook op het scherm verschijnen. Met name in de USA is A-CASI een standaardmethode voor gevoelige vragen geworden. Hoewel het gebruik van (A)-CASI survey-fouten sterk reduceert zijn deze zeker niet nul, er is nog veel ruimte voor verbetering van de accuraatheid. In de USA werkt een onderzoeksgroep aan de technische verbetering van A-CASI (Thornberry, Bhaskar, Krulewitch, Wesley, Hubbard, Adamson, 2000) zoals een

combinatie van A-CASI en touchscreens voor interviews bij speciale groepen over gevoelige onderwerpen. Maar aan de methodologische optimalisatie van de effectiviteit wordt nog te weinig aandacht besteed. Onderzoek naar het verder reduceren van de meetfouten bij efficiënte technieken zoals (A)-CASI is aan te bevelen.

In een meta-analyse van vergelijkend randomized response onderzoek dat gepubliceerd is tussen 1975 en 2000 blijkt dat de computer gestuurde zelfrapportage minder valide schatters oplevert dan de randomized responsemethode. Echter hierbij moet vermeld worden dat er maar weinig onderzoek is waarbij computer gestuurde zelfrapportages zijn vergeleken met randomized responsemethoden, zodat hier sprake kan zijn van een minder betrouwbaar resultaat (Lensvelt-Mulders, Hox en van der Heijden, opgestuurd). In plaats van beide methoden te vergelijken en tegen elkaar af te zetten is het ons inziens beter om het goede van beide methoden te verenigen in een computergestuurde vorm van randomized response zelfrapportage, CARR (zie volgende paragraaf).

3.2 'Locked box'-design

Een andere methode die speciaal is ontwikkeld om sensitieve vragen te stellen is de z.g.n. 'locked box' (gesloten doos) methode van McCochran (1982). Elke respondent krijgt hierbij een aantal kaarten voorgelegd met daarop beschreven een bepaald gedrag dat per kaart varieert in sensitiviteit. Daarbij wordt niet alleen het gedrag waarin de onderzoeker is geïnteresseerd beschreven maar ook algemeen voorkomende illegale of gevoelige gedragingen. In het geval van onderzoek naar regelovertreding kunnen de kaarten naast vragen over regelovertredend gedrag ook gedragingen als 'rijden door het rode licht' en 'onjuiste inkomsten opgeven bij de belastingdienst' bevatten. De respondent stopt de kaarten die een gedraging beschrijven waaraan hij of zij zich wel eens schuldig heeft gemaakt in doos A en kaarten die een gedraging beschrijven waaraan hij of zij zich niet schuldig heeft gemaakt in doos B. De interviewer kan niet zien welke kaart in welke doos gaat, pas later wordt de doos teruggekoppeld aan de respondent via een identificatiecode.

Elffers, van der Heijden en Hezemans (2003) hebben onderzoek gedaan naar de naleving van twee verschillende wetten waarbij de resultaten van de randomized response techniek werden vergeleken met die van de locked box methode. In onderstaande tabel kan men aflezen dat regelovertredingen vaker worden toegegeven door respondenten in de randomized response conditie dan door respondenten in de Locked box conditie.

Tabel 1: Regelovertreding

	Wet 1		Wet2	
	Locked box	RR	Locked box	RR
Aantal resp	345		323	
% 'nee answers'	6%	8%	3%	7%
Aantal valide antwoorden	325	317	314	299
% regelovertreding	11%	24%	4%	8%

Tabel 2 laat zien dat van 300 respondenten 223 bij beide methoden ontkennen de regels te hebben overtreden, 29 respondenten geven bij beide methoden toe de regels te hebben overtreden, en 44 respondenten geven in de randomized response conditie toe dat zij de regels hebben overtreden waar zij dit bij de locked box methode nog ontkende. De interpretatie van Elffers et al. (2003) is dat de randomized responsemethode meer valide schatters oplevert dan de locked box methode.

Tabel 2: Geschatte frequenties cross-classificatie methoden (wet 1)

		Randomized responsemethode	
		Toegeven regelovertreding	Ontkennen regelovertreding
Locked box methode	Toegeven regelovertreding	29.25	3.75
	Ontkennen regelovertreding	44.25	222.75

3.3 De unmatched count techniek

De unmatched count techniek wordt door sommigen gezien als een vorm van de randomized responsetechniek, maar werkt toch anders. De respondenten worden random in twee groepen verdeeld. Elke respondent krijgt een lijst voorgelegd met verschillende statements. Deze statements zijn zowel neutraal als gevoelig van aard. Subgroep 1 krijgt een lijst voorgelegd met 5 statements. Subgroep 2 krijgt een lijst voorgelegd met 6 statements. Het gedrag waarin de onderzoeker is geïnteresseerd wordt alleen voorgelegd aan subgroep 2. De respondenten moeten nu aangeven hoeveel statements op hen van toepassing zijn. Het verschil in gemiddelden tussen subgroep1 en subgroep2 is een maat voor het

voorkomen van het gevoelige gedrag. In een onderzoek naar winkeldiefstal door personeel gaf deze methode goede resultaten (Dalton, Wimbush en Daily, 1994).

De methode is zeer eenvoudig en nauwelijks invasief voor respondenten. Een nadeel van de methode is dat ook hier forse steekproeven nodig zijn en dat het de mogelijkheden om verbanden te leggen tussen het gevoelige gedrag en andere variabelen om het gedrag ook te verklaren, nog niet onderzocht zijn, iets dat bij randomized responseonderzoek al wel tot de mogelijkheden behoort.

4. Verder lezen

Wanneer u meer wil weten over de achtergronden van de methodologie van dataverzameling bij gevoelige onderwerpen in het algemeen en van randomized response technieken in het bijzonder dan raden we u de volgende literatuur aan:

4.1 Nederlands

De Leeuw, E. (1999). Vragen naar gevoelige informatie: een overzicht. *Pedagogische Studien*, 76, 2, 92-103.

De Leeuw, E. en Hox, J. (2000). Non-respons in surveys: een overzicht. *Kwalitatieve methoden*, 19, 31-53.

De Leeuw E., en G. Lensvelt-Mulders (2002). Vragen naar gevoelige informatie. *Facta*, 10, mei, 34-35.

De Leeuw, E., en G. Lensvelt-Mulders (2002). Beschermd door de dobbelsteen. *Facta*, 10, sept, 28-31.

Lensvelt-Mulders, G. (2003). Randomized Response Technieken als instrument voor het onderzoek van sociaal gevoelige onderwerpen. In: A.E Bronner, P. Dekker, J. C. Hoekstra, E. de Leeuw, Th. Poiesz, K. de Ruyter, en A. Smidts (Eds), *Ontwikkelingen in het marktonderzoek: jaarboek 2003*. Haarlem: de Vrieseborch.

Lensvelt-Mulders, G., en H. Boeije (2002). De implementatie van randomized response technieken in een CASI omgeving. Utrecht: UU, FSW, IOPS-UU.

Lensvelt-Mulders, G., van Gils, G.H.C., and van der Heijden, P.G.M. (2001). Een overzicht van schattingsmethoden voor de omvang van fraude. Utrecht: UU, FSW, IOPS-UU.

Van der Heijden, P. G. M, Hox, J. J., en Elffers (2002). Het meten van regelnaleving: een voorstudie. Utrecht: UU, FSW, IOPS-UU.

Van der Heijden, P. G. M. en de Leeuw, E. D. (2003). Gevoelige vragen, dobbelstenen, en de statistiek. *Facta*, 11.

4.2 Engels reviews

Chaudhuri, Arijit, & Rahul Mukerjee. 1988. *Randomized response: Theory and techniques*. New York: Marcel Dekker.

De Leeuw, E. Hox, J. en Kef, S. (in print). Computer assisted self interviewing tailored for special populations and topics. *Field Methods*.

Lee, Raymond M. 1993. *Doing research on sensitive topics*. London: Sage.

Lensvelt-Mulders, G., Hox, J. J., and van der Heijden (opgestuurd). Meta-analysis of randomized response research: 35 years of validation studies. *Sociological Methods and Research*.

- Lensvelt-Mulders, G., van der Heijden, P. G. M., and van Gils, G. (opgestuurd). Research designs for fraud studies: reviewing survey methods, audits, and capture-recapture methods.
- Rasinski, K. A., G. B. Willis, A. K. Baldwin, W. Yeh, & L. Lee. 1999. "Methods of data collection, perceptions of risks and losses, and motivation to give truthful answers to sensitive survey questions." *Applied Cognitive Psychology* 465-484.
- Richman, W. L., Kiesler, S., Weisband, S., Drasgow, F. (1999). A meta-analytic study of social desirability distortion in computer assisted questionnaires, traditional questionnaires and interviews. *Journal of Applied Psychology*, 84, 5, 754-775.
- Umesh, U.N. and Robert A. Peterson. 1991. "A Critical Evaluation of the Randomized Response Method. Applications, Validation, and Research Agenda." *Sociological Methods and Research* 20:104-138.

4.3 Engels methodebeschrijving

- Boeije, H. R. en Lensvelt-Mulders, G. (2002). Honest by chance: A qualitative interview study to clarify respondents' (non-)compliance with computer assisted randomized response. *Bulletin de Methodologie Sociologique* 75:24-39.
- Boruch, R. F. (1971). Assuring confidentiality of responses in social research: a note on strategies. *The American Sociologist*, 6(November), 308-311.
- Dalton, Dan R., James C. Wimbush, and Catherine M. Daily. 1994. "Using the unmatched count technique (UCT) to estimate base rates for sensitive behavior." *Personnel Psychology* 47:817-828
- Droitcour, Judith, Rachel A. Caspar, Michael L. Hubbard, Teresa L. Parsley, Wendy Visscher, and Trena M. Ezzati. 1991. "The item count technique as a method of indirect questioning: a review of its development and a case study application." Pp. 185-210 in *Measurement Errors in Surveys*, edited by P.P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz and S. Sudman. New York: Wiley.
- Greenberg, B. V., Abdul-Ela, A. A., Simmons, W. R., & Horvitz, D. G. (1969). The unrelated question randomized response model: Theoretical framework. *Journal of the American Statistical Association*, 66, 243-250.
- Kuk, Anthony Y. C. 1990. "Asking sensitive questions indirectly." *Biometrika* 77:436-438.
- Lensvelt-Mulders G., Hox, J. J. and Boeije, H. (2002). Complicated questionnaire procedures for sensitive topics: developing best practice procedures in a realistic setting. Proceedings of the conference on questionnaire development, evaluation and testing (QDET), Charleston, November 2002.
- Moors, J. J. A. (1971). Optimization of the unrelated question randomized response model. *Journal of the American Statistical Association*, 66(335), 627-629.

Musch, J., Bröder, A., en Klauer, K. C. (2001). Improving survey research on the World Wide Web using the randomized response technique. In: U. D. Reips, en M. Bosnjak (Eds), *Dimensions of Internet science*. Lengerich: Pabst.

Warner, Stanley L. 1965. "Randomized Response: A survey technique for eliminating evasive answer bias." *Journal of the American Statistical Association*, 60:63-69.

4.4 Engels algemeen vergelijkend onderzoek van 1990-2003

Armacost, Robert L., Jamshid C. Hosseini, Sara A. Morris and Kathleen A. Rehbein. 1991. "An empirical comparison of direct questioning, scenario, and randomized response methods for obtaining sensitive business." *Decision Science* 22:1073-1087.

Edgell, Stephen, E., Karen L. Duchan, and Samuel Himmelfarb. 1992. "An empirical test of the unrelated question randomized response technique." *Bulletin of the Psychonomic Society* 30:153-156.

Elffers, H., Van der Heijden, P., en Hezemans, M. (2003). Explaining regulatory (non)-compliance: A survey study of rule transgression for two Dutch instrumental laws, Applying the randomized response method. *Journal of Qualitative Criminology*.

Jobe, Jared B., William F. Pratt, Roger Tourangeau, Allison K. Baldwin and Kenneth A. Rasinski. 1997. "The Effects of Interview Mode on Sensitive Questions in a Fertility Survey." Pp. 311-330 in *Survey Measurement and Process Quality*, edited by L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz and D. Trewin. New York: Wiley.

Kerkvliet, Jan. 1994. "Cheating by Economics Students: A Comparison of Survey Results." *Journal-of-Economic-Education* 25:121-133.

Landsheer, Hans A., Peter van der Heijden, & Ger van Gils. 1999. "Trust and understanding, two psychological aspects of randomized response." *Quality and Quantity* 33:1-12.

Lensvelt-Mulders, G., Hox, J. J., and de Leeuw, E. 2002. Comparing the efficacy and efficiency of randomized response designs. Paper presented at the XV World Congress for Sociology, Brisbane, July 2002.

Nordlund, Sturla, Ingar Holme, and Steinar Tamsfoss. 1994. "Randomized response estimates for the purchase of smuggled liquor in Norway." *Addiction* 4:401-405.

Van der Heijden, Peter G. M., Ger van Gils, Jan Bouts, and Joop J. Hox. 1998. "A comparison of randomized response, CASAQ, and direct questioning; eliciting sensitive information in the context of social security fraud." *Kwantitatieve Methoden* 19:15-34.

Van der Heijden, Peter G. M., Ger van Gils, Jan Bouts, and Joop J. Hox. 2000. "A comparison of randomized response, CASI, and face to face direct questioning; eliciting sensitive information in the context of welfare and unemployment benefit." *Sociological Methods and Research* 28: 505-537.

- Williams, Bryan L. and Hoi Suen. 1994. "A methodological comparison of survey techniques in obtaining self-reports of condom-related behaviors." *Psychological Reports* 7: 1531-1537.
- Wimbush, Dan C. and Donald R. Dalton. 1997. "Base rate for employee theft: Convergence of multiple methods." *Journal of Applied Psychology* 82: 756-763.
- Wyatt, Gail E., Lawrence, A. Vodounon, and M. R Mickey. 1992. "The Wyatt sex history questionnaire: A structured interview for female sexual history taking." *Journal of Child Sexual Abuse* 1:51-68.

Bijlage A

Betrouwbaarheidsintervallen oneindige populaties

Tabel 1: Forced response ($p = 3/4$, $\theta = 1/6$)

Tabel 2: Kuk's kaartmethode ($\theta_1 = .8$, $\theta_2 = .2$)

Tabel 3: Directe vragen

De waarden in de cellen van de tabellen zijn $1.96\sqrt{\text{var}(\hat{\pi})}$.

Het 95% betrouwbaarheidsinterval is te berekenen als $\hat{\pi} \pm 1.96\sqrt{\text{var}(\hat{\pi})}$.

Bijlage A:

Tabel 1: Forced-responsmethode

pi	n																			
	100	200	300	400	500	600	700	800	900	1000	1100	1200	1300	1400	1500	1600	1700	1800	1900	2000
0.0	9.74	6.89	5.62	4.87	4.36	3.98	3.68	3.44	3.25	3.08	2.94	2.81	2.70	2.60	2.51	2.43	2.36	2.30	2.23	2.18
2.5	10.16	7.18	5.86	5.08	4.54	4.15	3.84	3.59	3.39	3.21	3.06	2.93	2.82	2.71	2.62	2.54	2.46	2.39	2.33	2.27
5.0	10.53	7.45	6.08	5.27	4.71	4.30	3.98	3.72	3.51	3.33	3.18	3.04	2.92	2.82	2.72	2.63	2.55	2.48	2.42	2.36
7.5	10.88	7.69	6.28	5.44	4.86	4.44	4.11	3.85	3.63	3.44	3.28	3.14	3.02	2.91	2.81	2.72	2.64	2.56	2.50	2.43
10.0	11.19	7.91	6.46	5.59	5.00	4.57	4.23	3.96	3.73	3.54	3.37	3.23	3.10	2.99	2.89	2.80	2.71	2.64	2.57	2.50
12.5	11.47	8.11	6.62	5.73	5.13	4.68	4.33	4.05	3.82	3.63	3.46	3.31	3.18	3.07	2.96	2.87	2.78	2.70	2.63	2.56
15.0	11.72	8.29	6.77	5.86	5.24	4.79	4.43	4.14	3.91	3.71	3.53	3.38	3.25	3.13	3.03	2.93	2.84	2.76	2.69	2.62
17.5	11.95	8.45	6.90	5.98	5.35	4.88	4.52	4.23	3.98	3.78	3.60	3.45	3.31	3.19	3.09	2.99	2.90	2.82	2.74	2.67
20.0	12.16	8.60	7.02	6.08	5.44	4.96	4.59	4.30	4.05	3.84	3.67	3.51	3.37	3.25	3.14	3.04	2.95	2.87	2.79	2.72
22.5	12.34	8.72	7.12	6.17	5.52	5.04	4.66	4.36	4.11	3.90	3.72	3.56	3.42	3.30	3.19	3.08	2.99	2.91	2.83	2.76
25.0	12.50	8.84	7.22	6.25	5.59	5.10	4.72	4.42	4.17	3.95	3.77	3.61	3.47	3.34	3.23	3.12	3.03	2.95	2.87	2.79
27.5	12.64	8.94	7.30	6.32	5.65	5.16	4.78	4.47	4.21	4.00	3.81	3.65	3.51	3.38	3.26	3.16	3.07	2.98	2.90	2.83
30.0	12.76	9.02	7.36	6.38	5.70	5.21	4.82	4.51	4.25	4.03	3.85	3.68	3.54	3.41	3.29	3.19	3.09	3.01	2.93	2.85
32.5	12.86	9.09	7.42	6.43	5.75	5.25	4.86	4.55	4.29	4.07	3.88	3.71	3.57	3.44	3.32	3.21	3.12	3.03	2.95	2.87
35.0	12.93	9.15	7.47	6.47	5.78	5.28	4.89	4.57	4.31	4.09	3.90	3.73	3.59	3.46	3.34	3.23	3.14	3.05	2.97	2.89
37.5	13.00	9.19	7.50	6.50	5.81	5.31	4.91	4.59	4.33	4.11	3.92	3.75	3.60	3.47	3.36	3.25	3.15	3.06	2.98	2.91
40.0	13.04	9.22	7.53	6.52	5.83	5.32	4.93	4.61	4.35	4.12	3.93	3.76	3.62	3.48	3.37	3.26	3.16	3.07	2.99	2.92
42.5	13.06	9.24	7.54	6.53	5.84	5.33	4.94	4.62	4.35	4.13	3.94	3.77	3.62	3.49	3.37	3.27	3.17	3.08	3.00	2.92
45.0	13.07	9.24	7.54	6.53	5.84	5.33	4.94	4.62	4.36	4.13	3.94	3.77	3.62	3.49	3.37	3.27	3.17	3.08	3.00	2.92
47.5	13.05	9.23	7.54	6.53	5.84	5.33	4.93	4.61	4.35	4.13	3.94	3.77	3.62	3.49	3.37	3.26	3.17	3.08	2.99	2.92
50.0	13.02	9.21	7.52	6.51	5.82	5.32	4.92	4.60	4.34	4.12	3.93	3.76	3.61	3.48	3.36	3.26	3.16	3.07	2.99	2.91
52.5	12.97	9.17	7.49	6.49	5.80	5.30	4.90	4.59	4.32	4.10	3.91	3.74	3.60	3.47	3.35	3.24	3.15	3.06	2.98	2.90
55.0	12.90	9.12	7.45	6.45	5.77	5.27	4.88	4.56	4.30	4.08	3.89	3.72	3.58	3.45	3.33	3.23	3.13	3.04	2.96	2.88
57.5	12.81	9.06	7.40	6.41	5.73	5.23	4.84	4.53	4.27	4.05	3.86	3.70	3.55	3.42	3.31	3.20	3.11	3.02	2.94	2.87
60.0	12.71	8.98	7.34	6.35	5.68	5.19	4.80	4.49	4.24	4.02	3.83	3.67	3.52	3.40	3.28	3.18	3.08	2.99	2.91	2.84
62.5	12.58	8.89	7.26	6.29	5.63	5.14	4.75	4.45	4.19	3.98	3.79	3.63	3.49	3.36	3.25	3.14	3.05	2.96	2.89	2.81
65.0	12.43	8.79	7.18	6.22	5.56	5.07	4.70	4.39	4.14	3.93	3.75	3.59	3.45	3.32	3.21	3.11	3.01	2.93	2.85	2.78
67.5	12.26	8.67	7.08	6.13	5.48	5.01	4.63	4.33	4.09	3.88	3.70	3.54	3.40	3.28	3.17	3.07	2.97	2.89	2.81	2.74
70.0	12.07	8.53	6.97	6.03	5.40	4.93	4.56	4.27	4.02	3.82	3.64	3.48	3.35	3.23	3.12	3.02	2.93	2.84	2.77	2.70
72.5	11.85	8.38	6.84	5.93	5.30	4.84	4.48	4.19	3.95	3.75	3.57	3.42	3.29	3.17	3.06	2.96	2.87	2.79	2.72	2.65
75.0	11.61	8.21	6.70	5.81	5.19	4.74	4.39	4.11	3.87	3.67	3.50	3.35	3.22	3.10	3.00	2.90	2.82	2.74	2.66	2.60
77.5	11.35	8.02	6.55	5.67	5.07	4.63	4.29	4.01	3.78	3.59	3.42	3.28	3.15	3.03	2.93	2.84	2.75	2.67	2.60	2.54
80.0	11.05	7.82	6.38	5.53	4.94	4.51	4.18	3.91	3.68	3.50	3.33	3.19	3.07	2.95	2.85	2.76	2.68	2.61	2.54	2.47
82.5	10.73	7.59	6.19	5.36	4.80	4.38	4.06	3.79	3.58	3.39	3.23	3.10	2.98	2.87	2.77	2.68	2.60	2.53	2.46	2.40
85.0	10.37	7.33	5.99	5.19	4.64	4.23	3.92	3.67	3.46	3.28	3.13	2.99	2.88	2.77	2.68	2.59	2.52	2.44	2.38	2.32
87.5	9.98	7.05	5.76	4.99	4.46	4.07	3.77	3.53	3.33	3.15	3.01	2.88	2.77	2.67	2.58	2.49	2.42	2.35	2.29	2.23
90.0	9.54	6.75	5.51	4.77	4.27	3.89	3.61	3.37	3.18	3.02	2.88	2.75	2.65	2.55	2.46	2.39	2.31	2.25	2.19	2.13
92.5	9.06	6.40	5.23	4.53	4.05	3.70	3.42	3.20	3.02	2.86	2.73	2.61	2.51	2.42	2.34	2.26	2.20	2.13	2.08	2.03
95.0	8.52	6.02	4.92	4.26	3.81	3.48	3.22	3.01	2.84	2.69	2.57	2.46	2.36	2.28	2.20	2.13	2.07	2.01	1.95	1.90
97.5	7.91	5.59	4.57	3.96	3.54	3.23	2.99	2.80	2.64	2.50	2.39	2.28	2.19	2.11	2.04	1.98	1.92	1.86	1.82	1.77
100.0	7.22	5.11	4.17	3.61	3.23	2.95	2.73	2.55	2.41	2.28	2.18	2.09	2.00	1.93	1.86	1.81	1.75	1.70	1.66	1.62

Bijlage A: (vervolg)

Tabel 2: Kuk's kaartmethode

pi	n																			
	100	200	300	400	500	600	700	800	900	1000	1100	1200	1300	1400	1500	1600	1700	1800	1900	2000
0.0	13.07	9.24	7.54	6.53	5.84	5.33	4.94	4.62	4.36	4.13	3.94	3.77	3.62	3.49	3.37	3.27	3.17	3.08	3.00	2.92
2.5	13.42	9.49	7.75	6.71	6.00	5.48	5.07	4.74	4.47	4.24	4.05	3.87	3.72	3.59	3.47	3.36	3.25	3.16	3.08	3.00
5.0	13.75	9.72	7.94	6.87	6.15	5.61	5.20	4.86	4.58	4.35	4.14	3.97	3.81	3.67	3.55	3.44	3.33	3.24	3.15	3.07
7.5	14.05	9.93	8.11	7.02	6.28	5.74	5.31	4.97	4.68	4.44	4.24	4.06	3.90	3.75	3.63	3.51	3.41	3.31	3.22	3.14
10.0	14.33	10.13	8.27	7.16	6.41	5.85	5.42	5.07	4.78	4.53	4.32	4.14	3.97	3.83	3.70	3.58	3.48	3.38	3.29	3.20
12.5	14.59	10.31	8.42	7.29	6.52	5.95	5.51	5.16	4.86	4.61	4.40	4.21	4.05	3.90	3.77	3.65	3.54	3.44	3.35	3.26
15.0	14.82	10.48	8.56	7.41	6.63	6.05	5.60	5.24	4.94	4.69	4.47	4.28	4.11	3.96	3.83	3.71	3.60	3.49	3.40	3.31
17.5	15.04	10.63	8.68	7.52	6.73	6.14	5.68	5.32	5.01	4.76	4.53	4.34	4.17	4.02	3.88	3.76	3.65	3.54	3.45	3.36
20.0	15.24	10.78	8.80	7.62	6.81	6.22	5.76	5.39	5.08	4.82	4.59	4.40	4.23	4.07	3.93	3.81	3.70	3.59	3.50	3.41
22.5	15.42	10.90	8.90	7.71	6.90	6.29	5.83	5.45	5.14	4.88	4.65	4.45	4.28	4.12	3.98	3.85	3.74	3.63	3.54	3.45
25.0	15.58	11.02	9.00	7.79	6.97	6.36	5.89	5.51	5.19	4.93	4.70	4.50	4.32	4.16	4.02	3.90	3.78	3.67	3.57	3.48
27.5	15.73	11.12	9.08	7.86	7.03	6.42	5.94	5.56	5.24	4.97	4.74	4.54	4.36	4.20	4.06	3.93	3.81	3.71	3.61	3.52
30.0	15.86	11.21	9.15	7.93	7.09	6.47	5.99	5.61	5.29	5.01	4.78	4.58	4.40	4.24	4.09	3.96	3.85	3.74	3.64	3.55
32.5	15.97	11.29	9.22	7.98	7.14	6.52	6.04	5.65	5.32	5.05	4.81	4.61	4.43	4.27	4.12	3.99	3.87	3.76	3.66	3.57
35.0	16.07	11.36	9.28	8.03	7.19	6.56	6.07	5.68	5.36	5.08	4.84	4.64	4.46	4.29	4.15	4.02	3.90	3.79	3.69	3.59
37.5	16.15	11.42	9.32	8.07	7.22	6.59	6.10	5.71	5.38	5.11	4.87	4.66	4.48	4.32	4.17	4.04	3.92	3.81	3.70	3.61
40.0	16.22	11.47	9.36	8.11	7.25	6.62	6.13	5.73	5.41	5.13	4.89	4.68	4.50	4.33	4.19	4.05	3.93	3.82	3.72	3.63
42.5	16.27	11.50	9.39	8.13	7.27	6.64	6.15	5.75	5.42	5.14	4.90	4.70	4.51	4.35	4.20	4.07	3.95	3.83	3.73	3.64
45.0	16.30	11.53	9.41	8.15	7.29	6.66	6.16	5.76	5.43	5.16	4.92	4.71	4.52	4.36	4.21	4.08	3.95	3.84	3.74	3.65
47.5	16.33	11.54	9.43	8.16	7.30	6.67	6.17	5.77	5.44	5.16	4.92	4.71	4.53	4.36	4.22	4.08	3.96	3.85	3.75	3.65
50.0	16.33	11.55	9.43	8.17	7.30	6.67	6.17	5.77	5.44	5.17	4.92	4.72	4.53	4.37	4.22	4.08	3.96	3.85	3.75	3.65
52.5	16.33	11.54	9.43	8.16	7.30	6.67	6.17	5.77	5.44	5.16	4.92	4.71	4.53	4.36	4.22	4.08	3.96	3.85	3.75	3.65
55.0	16.30	11.53	9.41	8.15	7.29	6.66	6.16	5.76	5.43	5.16	4.92	4.71	4.52	4.36	4.21	4.08	3.95	3.84	3.74	3.65
57.5	16.27	11.50	9.39	8.13	7.27	6.64	6.15	5.75	5.42	5.14	4.90	4.70	4.51	4.35	4.20	4.07	3.95	3.83	3.73	3.64
60.0	16.22	11.47	9.36	8.11	7.25	6.62	6.13	5.73	5.41	5.13	4.89	4.68	4.50	4.33	4.19	4.05	3.93	3.82	3.72	3.63
62.5	16.15	11.42	9.32	8.07	7.22	6.59	6.10	5.71	5.38	5.11	4.87	4.66	4.48	4.32	4.17	4.04	3.92	3.81	3.70	3.61
65.0	16.07	11.36	9.28	8.03	7.19	6.56	6.07	5.68	5.36	5.08	4.84	4.64	4.46	4.29	4.15	4.02	3.90	3.79	3.69	3.59
67.5	15.97	11.29	9.22	7.98	7.14	6.52	6.04	5.65	5.32	5.05	4.81	4.61	4.43	4.27	4.12	3.99	3.87	3.76	3.66	3.57
70.0	15.86	11.21	9.15	7.93	7.09	6.47	5.99	5.61	5.29	5.01	4.78	4.58	4.40	4.24	4.09	3.96	3.85	3.74	3.64	3.55
72.5	15.73	11.12	9.08	7.86	7.03	6.42	5.94	5.56	5.24	4.97	4.74	4.54	4.36	4.20	4.06	3.93	3.81	3.71	3.61	3.52
75.0	15.58	11.02	9.00	7.79	6.97	6.36	5.89	5.51	5.19	4.93	4.70	4.50	4.32	4.16	4.02	3.90	3.78	3.67	3.57	3.48
77.5	15.42	10.90	8.90	7.71	6.90	6.29	5.83	5.45	5.14	4.88	4.65	4.45	4.28	4.12	3.98	3.85	3.74	3.63	3.54	3.45
80.0	15.24	10.78	8.80	7.62	6.81	6.22	5.76	5.39	5.08	4.82	4.59	4.40	4.23	4.07	3.93	3.81	3.70	3.59	3.50	3.41
82.5	15.04	10.63	8.68	7.52	6.73	6.14	5.68	5.32	5.01	4.76	4.53	4.34	4.17	4.02	3.88	3.76	3.65	3.54	3.45	3.36
85.0	14.82	10.48	8.56	7.41	6.63	6.05	5.60	5.24	4.94	4.69	4.47	4.28	4.11	3.96	3.83	3.71	3.60	3.49	3.40	3.31
87.5	14.59	10.31	8.42	7.29	6.52	5.95	5.51	5.16	4.86	4.61	4.40	4.21	4.05	3.90	3.77	3.65	3.54	3.44	3.35	3.26
90.0	14.33	10.13	8.27	7.16	6.41	5.85	5.42	5.07	4.78	4.53	4.32	4.14	3.97	3.83	3.70	3.58	3.48	3.38	3.29	3.20
92.5	14.05	9.93	8.11	7.02	6.28	5.74	5.31	4.97	4.68	4.44	4.24	4.06	3.90	3.75	3.63	3.51	3.41	3.31	3.22	3.14
95.0	13.75	9.72	7.94	6.87	6.15	5.61	5.20	4.86	4.58	4.35	4.14	3.97	3.81	3.67	3.55	3.44	3.33	3.24	3.15	3.07
97.5	13.42	9.49	7.75	6.71	6.00	5.48	5.07	4.74	4.47	4.24	4.05	3.87	3.72	3.59	3.47	3.36	3.25	3.16	3.08	3.00
100.0	13.07	9.24	7.54	6.53	5.84	5.33	4.94	4.62	4.36	4.13	3.94	3.77	3.62	3.49	3.37	3.27	3.17	3.08	3.00	2.92

Bijlage A: (vervolg)

Tabel 3: Directe-vraagmethode





pi	n																			
	100	200	300	400	500	600	700	800	900	1000	1100	1200	1300	1400	1500	1600	1700	1800	1900	2000
0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2.5	3.06	2.16	1.77	1.53	1.37	1.25	1.16	1.08	1.02	0.97	0.92	0.88	0.85	0.82	0.79	0.77	0.74	0.72	0.70	0.68
5.0	4.27	3.02	2.47	2.14	1.91	1.74	1.61	1.51	1.42	1.35	1.29	1.23	1.18	1.14	1.10	1.07	1.04	1.01	0.98	0.96
7.5	5.16	3.65	2.98	2.58	2.31	2.11	1.95	1.83	1.72	1.63	1.56	1.49	1.43	1.38	1.33	1.29	1.25	1.22	1.18	1.15
10.0	5.88	4.16	3.39	2.94	2.63	2.40	2.22	2.08	1.96	1.86	1.77	1.70	1.63	1.57	1.52	1.47	1.43	1.39	1.35	1.31
12.5	6.48	4.58	3.74	3.24	2.90	2.65	2.45	2.29	2.16	2.05	1.95	1.87	1.80	1.73	1.67	1.62	1.57	1.53	1.49	1.45
15.0	7.00	4.95	4.04	3.50	3.13	2.86	2.65	2.47	2.33	2.21	2.11	2.02	1.94	1.87	1.81	1.75	1.70	1.65	1.61	1.56
17.5	7.45	5.27	4.30	3.72	3.33	3.04	2.81	2.63	2.48	2.36	2.25	2.15	2.07	1.99	1.92	1.86	1.81	1.76	1.71	1.67
20.0	7.84	5.54	4.53	3.92	3.51	3.20	2.96	2.77	2.61	2.48	2.36	2.26	2.17	2.10	2.02	1.96	1.90	1.85	1.80	1.75
22.5	8.18	5.79	4.73	4.09	3.66	3.34	3.09	2.89	2.73	2.59	2.47	2.36	2.27	2.19	2.11	2.05	1.99	1.93	1.88	1.83
25.0	8.49	6.00	4.90	4.24	3.80	3.46	3.21	3.00	2.83	2.68	2.56	2.45	2.35	2.27	2.19	2.12	2.06	2.00	1.95	1.90
27.5	8.75	6.19	5.05	4.38	3.91	3.57	3.31	3.09	2.92	2.77	2.64	2.53	2.43	2.34	2.26	2.19	2.12	2.06	2.01	1.96
30.0	8.98	6.35	5.19	4.49	4.02	3.67	3.39	3.18	2.99	2.84	2.71	2.59	2.49	2.40	2.32	2.25	2.18	2.12	2.06	2.01
32.5	9.18	6.49	5.30	4.59	4.11	3.75	3.47	3.25	3.06	2.90	2.77	2.65	2.55	2.45	2.37	2.30	2.23	2.16	2.11	2.05
35.0	9.35	6.61	5.40	4.67	4.18	3.82	3.53	3.31	3.12	2.96	2.82	2.70	2.59	2.50	2.41	2.34	2.27	2.20	2.14	2.09
37.5	9.49	6.71	5.48	4.74	4.24	3.87	3.59	3.35	3.16	3.00	2.86	2.74	2.63	2.54	2.45	2.37	2.30	2.24	2.18	2.12
40.0	9.60	6.79	5.54	4.80	4.29	3.92	3.63	3.39	3.20	3.04	2.90	2.77	2.66	2.57	2.48	2.40	2.33	2.26	2.20	2.15
42.5	9.69	6.85	5.59	4.84	4.33	3.96	3.66	3.43	3.23	3.06	2.92	2.80	2.69	2.59	2.50	2.42	2.35	2.28	2.22	2.17
45.0	9.75	6.89	5.63	4.88	4.36	3.98	3.69	3.45	3.25	3.08	2.94	2.81	2.70	2.61	2.52	2.44	2.36	2.30	2.24	2.18
47.5	9.79	6.92	5.65	4.89	4.38	4.00	3.70	3.46	3.26	3.10	2.95	2.83	2.71	2.62	2.53	2.45	2.37	2.31	2.25	2.19
50.0	9.80	6.93	5.66	4.90	4.38	4.00	3.70	3.46	3.27	3.10	2.95	2.83	2.72	2.62	2.53	2.45	2.38	2.31	2.25	2.19
52.5	9.79	6.92	5.65	4.89	4.38	4.00	3.70	3.46	3.26	3.10	2.95	2.83	2.71	2.62	2.53	2.45	2.37	2.31	2.25	2.19
55.0	9.75	6.89	5.63	4.88	4.36	3.98	3.69	3.45	3.25	3.08	2.94	2.81	2.70	2.61	2.52	2.44	2.36	2.30	2.24	2.18
57.5	9.69	6.85	5.59	4.84	4.33	3.96	3.66	3.43	3.23	3.06	2.92	2.80	2.69	2.59	2.50	2.42	2.35	2.28	2.22	2.17
60.0	9.60	6.79	5.54	4.80	4.29	3.92	3.63	3.39	3.20	3.04	2.90	2.77	2.66	2.57	2.48	2.40	2.33	2.26	2.20	2.15
62.5	9.49	6.71	5.48	4.74	4.24	3.87	3.59	3.35	3.16	3.00	2.86	2.74	2.63	2.54	2.45	2.37	2.30	2.24	2.18	2.12
65.0	9.35	6.61	5.40	4.67	4.18	3.82	3.53	3.31	3.12	2.96	2.82	2.70	2.59	2.50	2.41	2.34	2.27	2.20	2.14	2.09
67.5	9.18	6.49	5.30	4.59	4.11	3.75	3.47	3.25	3.06	2.90	2.77	2.65	2.55	2.45	2.37	2.30	2.23	2.16	2.11	2.05
70.0	8.98	6.35	5.19	4.49	4.02	3.67	3.39	3.18	2.99	2.84	2.71	2.59	2.49	2.40	2.32	2.25	2.18	2.12	2.06	2.01
72.5	8.75	6.19	5.05	4.38	3.91	3.57	3.31	3.09	2.92	2.77	2.64	2.53	2.43	2.34	2.26	2.19	2.12	2.06	2.01	1.96
75.0	8.49	6.00	4.90	4.24	3.80	3.46	3.21	3.00	2.83	2.68	2.56	2.45	2.35	2.27	2.19	2.12	2.06	2.00	1.95	1.90
77.5	8.18	5.79	4.73	4.09	3.66	3.34	3.09	2.89	2.73	2.59	2.47	2.36	2.27	2.19	2.11	2.05	1.99	1.93	1.88	1.83
80.0	7.84	5.54	4.53	3.92	3.51	3.20	2.96	2.77	2.61	2.48	2.36	2.26	2.17	2.10	2.02	1.96	1.90	1.85	1.80	1.75
82.5	7.45	5.27	4.30	3.72	3.33	3.04	2.81	2.63	2.48	2.36	2.25	2.15	2.07	1.99	1.92	1.86	1.81	1.76	1.71	1.67
85.0	7.00	4.95	4.04	3.50	3.13	2.86	2.65	2.47	2.33	2.21	2.11	2.02	1.94	1.87	1.81	1.75	1.70	1.65	1.61	1.56
87.5	6.48	4.58	3.74	3.24	2.90	2.65	2.45	2.29	2.16	2.05	1.95	1.87	1.80	1.73	1.67	1.62	1.57	1.53	1.49	1.45
90.0	5.88	4.16	3.39	2.94	2.63	2.40	2.22	2.08	1.96	1.86	1.77	1.70	1.63	1.57	1.52	1.47	1.43	1.39	1.35	1.31
92.5	5.16	3.65	2.98	2.58	2.31	2.11	1.95	1.83	1.72	1.63	1.56	1.49	1.43	1.38	1.33	1.29	1.25	1.22	1.18	1.15
95.0	4.27	3.02	2.47	2.14	1.91	1.74	1.61	1.51	1.42	1.35	1.29	1.23	1.18	1.14	1.10	1.07	1.04	1.01	0.98	0.96
97.5	3.06	2.16	1.77	1.53	1.37	1.25	1.16	1.08	1.02	0.97	0.92	0.88	0.85	0.82	0.79	0.77	0.74	0.72	0.70	0.68
100.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Bijlage B: Betrouwbaarheidsintervallen eindige populaties

Celwaarden zijn $1.96\sqrt{\text{var}(\hat{\pi})}$ voor de forced-responsemethode bij verschillende combinaties van n , N en $\hat{\pi}$ (95% betrouwbaarheidsinterval is geklikt aan $\hat{\pi} \pm \text{celwaarde}$)

$\hat{\pi}$	n	N										
		200	400	600	800	1000	1200	1400	1600	1800	2000	oneindig
5%	50	14.59	14.75	14.80	14.82	14.84	14.85	14.85	14.86	14.86	14.87	14.90
	100	10.09	10.31	10.39	10.43	10.45	10.47	10.47	10.49	10.49	10.49	10.53
	200	6.80	7.13	7.23	7.29	8.23	7.35	7.37	7.37	7.39	7.39	7.45
	300	-	5.70	5.82	5.90	5.94	5.96	5.98	5.98	6.00	6.00	6.08
	400	-	4.82	4.98	5.04	5.10	5.12	5.14	5.15	5.17	5.17	5.27
	500	-	-	4.37	4.47	4.51	4.55	4.57	4.59	4.61	4.61	4.70
	600	-	-	3.94	4.02	4.08	4.12	4.16	4.17	4.17	4.19	4.29
	700	-	-	-	3.68	3.74	3.78	3.82	3.84	3.86	3.86	3.98
	800	-	-	-	3.41	3.47	3.51	3.55	3.57	3.59	3.61	3.72
10%	50	15.27	15.55	15.64	15.69	15.71	15.73	15.74	15.75	15.76	15.77	15.82
	100	10.39	10.80	10.92	11.00	11.03	11.05	11.07	11.09	11.09	11.11	11.19
	200	6.72	7.35	7.55	7.62	7.68	7.72	7.76	7.78	7.78	7.80	7.92
	300	-	5.74	6.00	6.12	6.19	6.23	6.27	6.29	6.31	6.33	6.47
	400	-	4.76	5.06	5.19	5.27	5.33	5.37	5.39	5.41	5.43	5.59
	500	-	-	4.39	4.55	4.65	4.70	4.74	4.78	4.80	4.82	5.00
	600	-	-	3.88	4.06	4.17	4.23	4.29	4.33	4.35	4.37	4.57
	700	-	-	-	3.68	3.80	3.88	3.92	3.96	4.00	4.02	4.23
	800	-	-	-	3.37	3.49	3.57	3.63	3.67	3.70	3.72	3.96
20%	50	16.29	16.75	16.90	16.97	17.02	17.05	17.07	17.08	17.09	17.10	17.19
	100	10.82	11.51	11.72	11.84	11.90	11.94	11.98	12.00	12.01	12.03	12.15
	200	6.57	7.64	7.98	8.13	8.23	8.29	8.33	8.37	8.39	8.41	8.60
	300	-	5.82	6.25	6.45	6.57	6.64	6.70	6.74	6.76	6.80	7.02
	400	-	4.65	5.17	5.41	5.55	5.64	5.70	5.76	5.78	5.82	6.08
	500	-	-	4.39	4.68	4.84	4.94	5.02	5.08	5.12	5.15	5.43
	600	-	-	3.80	4.12	4.29	4.41	4.51	4.57	4.61	4.65	4.96
	700	-	-	-	3.67	3.86	4.00	4.10	4.16	4.21	4.25	4.59
	800	-	-	-	3.29	3.51	3.65	3.74	3.82	3.88	3.92	4.29
40%	50	17.16	17.81	18.02	18.13	18.19	18.23	18.26	18.28	18.30	18.31	18.44
	100	11.13	12.11	12.43	12.58	12.68	12.74	12.78	12.82	12.84	12.86	13.03
	200	6.23	7.88	8.35	8.57	8.70	8.80	8.86	8.90	8.94	8.96	9.21
	300	-	5.80	6.43	6.72	6.88	7.00	7.08	7.13	7.17	7.21	7.53
	400	-	4.41	5.21	5.57	5.76	5.90	6.00	6.06	6.12	6.15	6.53
	500	-	-	4.31	4.74	4.98	5.14	5.23	5.31	5.37	5.43	5.82
	600	-	-	3.61	4.10	4.37	4.55	4.66	4.74	4.82	4.86	5.33
	700	-	-	-	3.57	3.88	4.08	4.21	4.31	4.37	4.43	4.92
	800	-	-	-	3.12	3.47	3.68	3.82	3.94	4.02	4.08	4.61

LEGENDA: Omvang en kwalificatie betrouwbaarheidsintervallen:

	Slecht:	$95\% \text{ betrouwbaarheidsinterval} > \hat{\pi} \pm 100\% * \hat{\pi}$
	Matig:	$\hat{\pi} \pm 100\% * \hat{\pi} > 95\% \text{ betrouwbaarheidsinterval} > \hat{\pi} \pm 80\% * \hat{\pi}$
	Redelijk:	$\hat{\pi} \pm 80\% * \hat{\pi} > 95\% \text{ betrouwbaarheidsinterval} > \hat{\pi} \pm 60\% * \hat{\pi}$
	Goed:	$\hat{\pi} \pm 60\% * \hat{\pi} > 95\% \text{ betrouwbaarheidsinterval}$