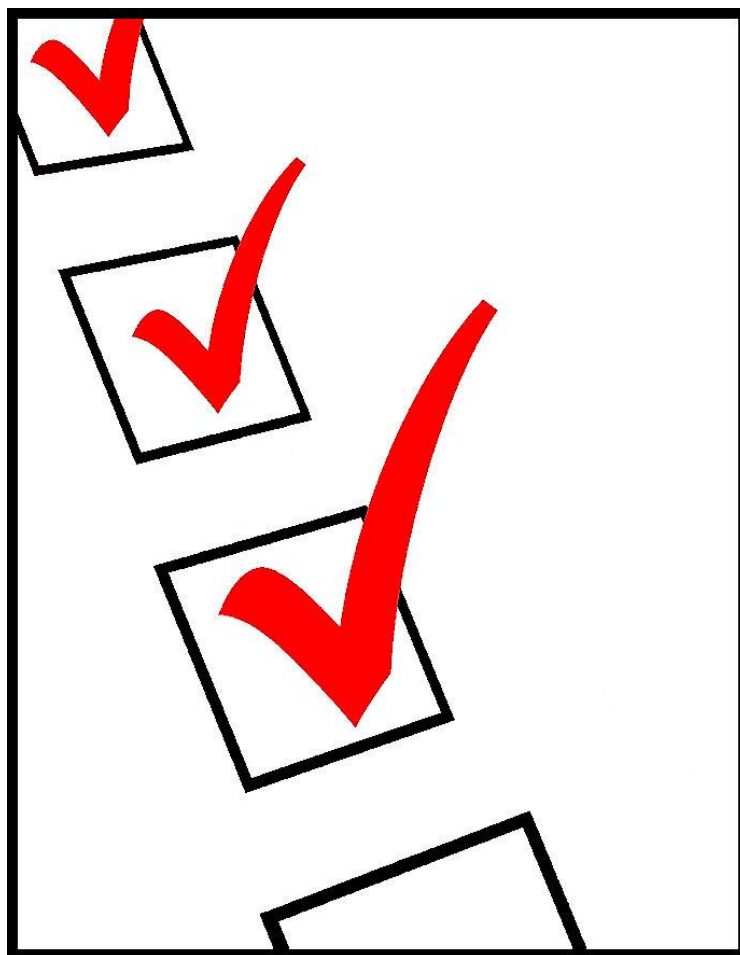


Checklist voor peilingen



Jelke Bethlehem

Versie 2.0 (6 juli 2010)

Een checklist voor peilingen

Inleiding

Er wordt in Nederland heel veel gepeild. Dat is vooral te merken in de periode voor de Tweede Kamerverkiezingen. In die campagnetijd volgen de politieke peilingen elkaar in hoog tempo op. Maar ook buiten de verkiezingen om wordt de mening van 'de Nederlander' steeds vaker gevraagd over allerlei onderwerpen.

Peilingen, enquêtes, surveys ... Het zijn verschillende namen voor hetzelfde type onderzoek. Aan een steekproef van personen uit een populatie wordt een vragenlijst voorgelegd. Die vragen kunnen gaan over feitelijke zaken, gedrag en meningen. Het is dan de bedoeling om met de verkregen antwoorden uitspraken te doen over de gehele populatie. Dat kan, maar dan moet het onderzoek wel op wetenschappelijk verantwoorde wijze worden opgezet en uitgevoerd

Vooraf het Internet is er de oorzaak van dat het aantal peilingen sterk is toegenomen. Internet maakt het mogelijk eenvoudig, snel en goedkoop bij heel veel mensen informatie te verzamelen. De vraag is echter of al die peilingen wel een goed beeld geven van de werkelijkheid.

Er wordt dus steeds meer onderzoek gedaan. Daar zitten goede en slechte onderzoeken bij. Het is niet eenvoudig voor gebruikers van de uitkomsten van peilingen (journalisten, bestuurders, beleidsmakers) om op simpele wijze het kaf van het koren te scheiden. Deze checklist is een handreiking aan hen. Door de vragen in de checklist één voor één af te lopen, kan een eerste indruk worden verkregen van de kwaliteit van een onderzoek. Lijkt de kwaliteit goed te zijn, dan verdienen de uitkomsten van de peiling nadere aandacht. Roept het doorlopen van de checklist veel vragen op over de kwaliteit van een onderzoek, dan kan worden besloten er verder geen aandacht aan te besteden, of bij de beschrijving van het onderzoek kan worden vermeld welke problemen er mogelijk kunnen zijn.

Er is geprobeerd de checklist simpel te houden. Daardoor heeft hij een uitgesproken karakter gekregen: een aanpak is goed of fout. Uiteraard is de werkelijkheid vaak complexer, en dan zal het eindoordeel ook wat genuanceerder moeten zijn.

Hieronder worden de vragen in de checklist nog wat nader toegelicht. Uit de uitleg zal blijken waarom beantwoording van de vraag belangrijk is.

1. Wie is de opdrachtgever?

Het is van belang te weten wie de opdrachtgever en/of financier is van het onderzoek. Het komt voor dat die er belang bij heeft dat de uitkomsten een bepaalde richting uitgaan. Zo worden er regelmatig persberichten verspreid met de uitkomsten van onderzoek waaruit zou moeten blijken dat bepaalde producten of diensten erg goed zijn. Die persberichten blijken dan afkomstig te zijn van bedrijven die deze producten of diensten leveren. Het gaat dan louter om de publiciteit en niet om objectief onderzoek.

De BBC kent de richtlijn dat de opdrachtgever van de peiling de uitkomsten van een opiniepeiling niet in een uitzending mag komen toelichten of interpreteren. De omroep dient zich te beperken tot het slechts melden van de uitkomsten. Ook moet de opdrachtgever worden vermeld. Uiteraard moet dan wel bekend zijn wie de opdrachtgever is.

2. Is er een onderzoeksverantwoording beschikbaar?

De onderzoeksverantwoording is een rapport dat precies beschrijft hoe het onderzoek is opgezet en uitgevoerd. Het rapport moet voldoende informatie bevatten om te kunnen vaststellen of het om een wetenschappelijk verantwoorde onderzoek gaat.

De onderzoeksverantwoording moet minimaal de volgende aspecten beschrijven:

- De naam van de opdrachtgever;
- De naam van de organisatie die de peiling heeft uitgevoerd;
- De doelpopulatie. Dat is de groep waarop de uitkomsten van het onderzoek betrekking heeft;
- De volledige vragenlijst. Duidelijk moet worden gemaakt of de vragenlijst ook van tevoren is getest;
- Het gebruikte steekproefkader. Dit is de lijst waarin voor ieder lid van de doelpopulatie de contactgegevens staan (adres, telefoonnummer, e-mailadres, e.d.);
- De wijze waarop de steekproef is getrokken. Duidelijk moet worden aangegeven of er is geloot en hoe er is geloot;
- De omvang de bruto steekproef. Dit is de omvang van de steekproef zoals hij is getrokken uit het steekproefkader;
- De omvang van de respons (netto steekproef). Dit is het aantal respondenten.
- Het percentage respons ($100 \times \text{Response} / \text{Bruto steekproef}$).
- De wijze waarop de respons eventueel is gecorrigeerd voor non-respons (en eventuele andere selectie-effecten). Minimaal moeten de gebruikte weegvariabelen worden genoemd. Bij voorkeur moet ook worden uitgelegd hoe die variabelen zijn gebruikt in de berekening van de gewichten.
- De omvang van de onzekerheidsmarges. Merk op dat deze alleen kunnen worden berekend als de steekproef is geloot, en de non-respons niet tot vertekening leidt. Ingeval van aanzienlijke non-respons of zelfselectie kunnen de afwijkingen nog veel groter zijn. Dit moet worden gemeld.

3. Wat is de doelpopulatie

De doelpopulatie is de groep van mensen waarop de uitkomsten van het onderzoek betrekking heeft. De doelpopulatie moet duidelijk worden omschreven.

Problemen kunnen ontstaan als het steekproefkader niet de hele populatie afdekt. In feite wordt de steekproef dan uit een andere, kleinere groep getrokken. De consequentie daarvan is dat de uitkomsten dan ook betrekking op die kleinere groep en niet op de oorspronkelijke populatie.

Voorbeeld: de doelpopulatie is gedefinieerd als alle Nederlanders van 18 jaar en ouder, terwijl de steekproef wordt getrokken uit bezitters van een Internetverbinding thuis. De

conclusies van het onderzoek hebben dan alleen betrekking op de Internetbezitters van 18 jaar en ouder.

4. Hoe zit het met de kwaliteit van de vragenlijst?

De vragenlijst is van cruciaal belang. Door te manipuleren met de teksten van de vragen en de volgorde van de vragen is het mogelijk de uitkomsten van het onderzoek te beïnvloeden. In een goede vragenlijst zijn de vragen objectief en begrijpelijk geformuleerd.

In ieder geval moeten de volgende valkuilen worden vermeden:

- Onbegrijpelijke vragen. Vragen kunnen onbegrijpelijk worden voor de respondenten door jargon of woordgebruik dat hen niet bekend is, of door lange, vage of ingewikkelde zinsconstructies. Voorbeeld: *Bent u tevreden over de recreatieve voorziening in uw woonplaats?*
- Vragen die op verschillende manieren kunnen worden geïnterpreteerd. Voorbeeld: *Wanneer ging u van school af?* Gaat het hier om een datum, leeftijd of andere gebeurtenis (toen ik in verwachting was)?
- Suggestieve en sturende vragen. Voorbeeld: *Vindt u ook niet dat ... ?*
- Vragen waarin meer dan één vraag wordt gesteld. Voorbeeld: Heeft u deze baan voor korte tijd om iets bij te verdienen?
- Vragen die een ontkenning of dubbele ontkenning bevatten. Voorbeeld: *Gebruikt u liever niet een niet-medicinale shampoo?*
- Vragen die teveel een beroep doen op de herinnering. Dit levert vooral problemen op als het om een relatief lange periode gaat en gevraagd wordt naar gebeurtenissen die niet heel belangrijk zijn en regelmatig voorkomen. Voorbeeld: *Hoe vaak heeft u in het afgelopen jaar een bezoek gebracht aan de supermarkt?.*

Voor een goed onderzoek is het noodzakelijk dat de vragenlijst wordt getest voordat hij echt gebruikt gaat worden voor het onderzoek.

5. Hoe is de steekproef getrokken?

Om op verantwoorde wijze conclusies te kunnen trekken over de doelpopulatie, moet de steekproef door loting worden getrokken. Het moet een kanssteekproef zijn. Iedereen in de populatie moet een positieve kans hebben gehad om in de steekproef te komen. De trekkingskansen moeten bekend zijn.

De eenvoudigste wijze van trekken is die waarbij iedereen dezelfde kans heeft gehad om in de steekproef te komen. Dan geldt het principe dat het percentage (of gemiddelde) in de steekproef een goede schatter is voor het percentage (of gemiddelde) in de populatie.

Het is denkbaar om de steekproef met ongelijke kansen te trekken. Het berekenen van schattingen voor de doelpopulatie is dan wat ingewikkelder, omdat gecorrigeerd moet worden voor die ongelijke kansen. Een voorbeeld is een steekproef van personen die wordt verkregen door eerst adressen met gelijke kansen te loten en vervolgens op elk geselecteerd adres een willekeurig persoon te kiezen. Personen in grote gezinnen hebben dan een kleinere trekkingskans dan personen in kleine gezinnen.

De steekproef wordt getrokken uit een steekproefkader. Dit is de (elektronische) lijst waarin voor ieder lid van de doelpopulatie de contactgegevens staan (adres, telefoonnummer, e-mailadres, e.d.). Als het steekproefkader niet de gehele doelpopulatie bestrijkt, dan hebben de conclusies van het onderzoek alleen betrekking op het deel van de doelpopulatie dat via het steekproefkader kan worden bereikt.

Als de steekproef niet is verkregen door loting, dan zijn de trekkingskansen onbekend en kunnen nooit goede schattingen voor populatiepercentages (en populatiegemiddelden) worden berekend. Dit is bijvoorbeeld het geval bij peilingen via het Internet waarbij de steekproef is geworven door zelfselectie.

6. Hoe groot is de steekproef?

Als de steekproef door loting is verkregen, dan kan de nauwkeurigheid van de schattingen worden berekend. Daarbij geldt de eenvoudige regel dat de nauwkeurigheid toeneemt als de steekproef groter is.

De nauwkeurigheid wordt meestal weergegeven in de vorm van een onzekerheidsmarge. Die onzekerheidsmarge geeft aan hoeveel de schatting maximaal kan afwijken van de werkelijke waarde in de doelpopulatie. Tabel 1 bevat voor verschillende steekproefgroottes en een reeks van steekproefpercentages de bijbehorende onzekerheidsmarge.

Stel dat uit een peiling onder een steekproef van 500 personen blijkt dat 40% van de respondenten voorstander is van een bepaalde maatregel is. Bij die 500 en die 40% hoort een marge van 4,3%. Het percentage voorstanders in de doelpopulatie zal dus liggen tussen $40 - 4,3 = 35,7\%$ en de $40 + 4,3 = 44,3\%$.

Stel dat in een peiling onder 1000 personen wordt gevraagd naar de politieke voorkeur. Uit de peiling blijkt dat 20% op een bepaalde partij gaat stemmen. Een maand later wordt een nieuwe peiling uitgevoerd, ook onder 1000 personen. Nu blijkt 22% op die partij te gaan stemmen. Kun je nu zeggen dat de aanhang van die partij is toegenomen? Nee, want beide percentage hebben een onzekerheidsmarge van 2,5%. Die marge is groter dan het verschil tussen de percentage stemmers ($22 - 20 = 2\%$). Het verschil kan dus net zo goed veroorzaakt door de 'ruis' in de steekproef.

Figuur 1. Onzekerheidsmarges

Percentage	Omvang van de steekproef					
	100	200	500	1000	2000	5000
10	5,9	4,2	2,6	1,9	1,3	0,8
20	7,9	5,6	3,5	2,5	1,8	1,1
30	9,0	6,4	4,0	2,8	2,0	1,3
40	9,7	6,8	4,3	3,0	2,1	1,4
50	9,8	6,9	4,4	3,1	2,2	1,4
60	9,7	6,8	4,3	3,0	2,1	1,4
70	9,0	6,4	4,0	2,8	2,0	1,3
80	7,9	5,6	3,5	2,5	1,8	1,1
90	5,9	4,2	2,6	1,9	1,3	0,8

7. Hoe groot is de non-respons?

Non-respons is het verschijnsel dat in de steekproef getrokken personen niet de gewenste gegevens verstrekken. Het lukt niet contact met ze te leggen (bijvoorbeeld omdat ze zelden thuis zijn), ze zijn niet in staat om de vragen te beantwoorden (door bijvoorbeeld ziekte of taalproblemen) of ze weigeren mee te doen. Non-respons leidt er heel vaak toe dat bepaalde groepen mensen over- of ondervertegenwoordigd zijn in de steekproef. Daardoor is de gerealiseerde steekproef niet meer representatief. De omvang van de vertekening in de uitkomsten wordt door drie factoren bepaald:

- Hoeveel mensen doen niet mee? Een groter percentage non-respons groter leidt tot een grotere afwijking in de uitkomsten. Is de respons bij een peiling laag, dan is het risico groot dat er ernstige afwijkingen zijn in de uitkomsten.
- Zijn non-respondenten anders? Bijvoorbeeld: bij verkiezingsonderzoek blijkt vaak een sterke samenhang tussen respons en stemgedrag. Mensen die niet meedoen aan het onderzoek gaan ook niet stemmen. En respondenten gaan juist wel stemmen. Dit leidt ertoe dat stemmers zijn oververtegenwoordigd in de peiling. Als gevolg wordt de opkomst te hoog geschat.
- Doen sommige mensen vaker aan onderzoek mee dan anderen? Als er mensen zijn met hoge responskansen (ze doen vaak mee aan dit soort onderzoek) en mensen met lage responskansen (ze doen bijna nooit mee aan dit soort onderzoek), dan verhoogt dit de vertekening in de uitkomsten. Als iedereen even geneigd is om aan het onderzoek mee te doen (hun responskansen zijn gelijk), dan leidt non-respons niet tot een vertekening.

Het is meestal niet mogelijk om vast te stellen hoe groot de vertekening is. Dat kan alleen als bekend is wat de antwoorden van de non-respondenten op de vragen zouden zijn. Omdat het echter om non-respondenten gaat, zijn die antwoorden dus onbekend. Het is wel mogelijk om uit te rekenen hoe groot de afwijking ten gevolge van non-respons maximaal kan zijn.

Stel eens dat slechts 40% van de mensen in de steekproef respondeert in een verkiezingsonderzoek. Van die respondenten zegt 60% dat ze zullen gaan stemmen. Als 40% respondeert, dan respondeert 60% dus niet. Stel dat in een extreme situatie al die non-respondenten niet gaan stemmen. Dan wordt het percentage stemmers in de hele steekproef

$$0,40 \times 60\% + 0,60 \times 0\% = 24\%.$$

In de andere extreme situatie gaan alle non-respondenten wel stemmen. Dan wordt het percentage stemmers in de hele steekproef

$$0,40 \times 60\% + 0,60 \times 100\% = 84\%.$$

Dus het in de respons gevonden percentage stemmers van 40% had ook 24% kunnen zijn en het had ook 84% kunnen zijn. Het werkelijk percentage stemmers zal ergens tussen 24% en 84% liggen. Een dergelijke onzekerheidsmarge maakt onderzoek

onbruikbaar. Gelukkig doen die extreme situaties zich in de praktijk niet vaak voor, maar het effect van non-respons kan groot zijn.

Tabel 2 bevat de grenzen voor de werkelijke percentages in de steekproef voor een reeks aan verschillende responspercentages. Duidelijk is te zien dat de omvang van de maximale afwijking kleiner wordt naarmate het percentage respons groter is.

Figuur 2. Bandbreedte ten gevolge van non-respons

Gevonden percentage	Percentage respons			
	20	40	60	80
10	2 - 82	4 - 64	6 - 46	8 - 28
20	4 - 84	8 - 68	12 - 52	16 - 36
30	6 - 86	12 - 72	18 - 58	24 - 44
40	8 - 88	16 - 76	24 - 64	32 - 52
50	10 - 90	20 - 80	30 - 70	40 - 60
60	12 - 92	24 - 84	36 - 76	48 - 68
70	14 - 94	28 - 88	42 - 82	56 - 76
80	16 - 96	32 - 92	48 - 88	64 - 84
90	18 - 98	36 - 96	54 - 94	72 - 92

8. Is er voor non-respons gecorrigeerd?

Het is gebruikelijk om te corrigeren voor non-respons door wegen: aan elke respondent een gewicht wordt toegekend. Die gewichten worden zo berekend dat ze corrigeren voor de over- of ondervertegenwoordiging van groepen in de respons.

Een voorbeeld: uit een peiling blijkt dat de respondenten voor 60% uit mannen bestaan en voor 40% uit vrouwen. Dat is niet goed want de Nederlandse bevolking bestaat voor 49,5% uit mannen en voor 50,5% uit vrouwen. Kennelijk hebben de mannen beter gerepondeerd in de peiling wat ze zijn oververtegenwoordigd. Om hiervoor te corrigeren, krijgt elke responderende man een gewicht van

$$49,5 / 60 = 0,825.$$

Dit betekent dat elke man nog maar meetelt voor 0,825 in plaats van 1. Het gewicht is kleiner dan 1 omdat er teveel mannen waren in de respons. Elk vrouw krijgt een gewicht van

$$50,5 / 40 = 1,263.$$

Dus elke vrouw telt mee voor 1,263 in plaats van 1. Het gewicht is groter dan 1 omdat er te weinig vrouwen waren in de steekproef.

Het toekennen van gewichten aan mannen en vrouwen leidt ertoe dat de respons representatief wordt met betrekking tot het geslacht. Die gewichten konden worden uitgerekend omdat de juiste percentages in de populatie bekend waren. Het idee achter wegen is nu om de respons representatief te maken met betrekking tot zoveel mogelijk kenmerken. Noodgedwongen kunnen dat alleen kenmerken zijn die in de peiling zijn

gemeten waarvoor de verdeling in de populatie bekend is. Voorbeelden van veel gebruikte kenmerken zijn geslacht, leeftijd, burgerlijke staat en provincie waar men woont.

De hoop is dat als de respons representatief is gemaakt met betrekking tot zoveel mogelijk variabelen, hij misschien ook wel representatief wordt met betrekking tot de verschijnselen die worden onderzocht. Daarmee zou dan de vertekening in de uitkomsten verdwijnen.

Helaas is niet elke weging zonder meer effectief. Wegen werkt alleen goed om een vertekening ten gevolge van non-respons te verminderen als voldaan is aan twee voorwaarden:

- Er is een sterke samenhang tussen het verschijnsel dat wordt onderzocht en de kenmerken die worden gebruikt voor wegen;
- Er is een sterke samenhang tussen het responsgedrag en de kenmerken die worden gebruikt voor wegen.

Soms heeft wegen veel effect, bijvoorbeeld als bij een onderzoek naar tijdbesteding wordt gewogen naar het hebben van een baan en de aanwezigheid van kinderen in het gezin. Die factoren hebben ook een invloed op de tijdbesteding. Maar het wegen met leeftijd en geslacht in een onderzoek naar opinies is vaak zinloos, omdat die kenmerken opinies niet bepalen. In alle gevallen is het van belang om zoveel mogelijk na te gaan welke groepen ondervertegenwoordigd zijn, wat dat voor effect op de resultaten kan hebben en of wegen de uitkomsten van het onderzoek beïnvloed.

9. Zijn er onzekerheidsmarges gegeven?

De uitkomsten van een peiling zijn slechts schattingen van wat er in de populatie aan de hand is. Het is daarom onrealistisch te veronderstellen dat die uitkomsten exact overeenkomen met de werkelijke cijfers voor de hele populatie. Zelfs in het ideale geval van een nette kanssteekproef zonder non-respons is er een afwijking tussen schatting en werkelijke waarde. Het is echter wel mogelijk om uit te rekenen hoe groot die afwijking maximaal kan zijn. Dat is de onzekerheidsmarge. Voor het bepalen van de onzekerheidsmarge kan tabel 1 worden gebruikt.

Voorbeeld: voor een peiling wordt een steekproef van 500 personen geloot. Iedereen doet mee. Er is geen non-respons. Van de respondenten zegt 60% voorstander te zijn van een bepaalde maatregel. De bijbehorende onzekerheidsmarge is 4,3%. Dan zal het percentage voorstanders in de populatie liggen tussen de $60 - 4,3 = 55,7\%$ en de $60 + 4,3 = 64,3\%$.

Zoals eerder is gebleken kunnen de onzekerheidsmarges als gevolg van non-respons veel groter zijn dan die gebaseerd op de steekproefomvang. Het gaat hier dus vooral om een ondergrens. Als veranderingen in de tijd kleiner zijn dan de onzekerheidsmarges veroorzaakt door de steekproefomvang, dan is er geen sprake van 'echte' verschillen tussen peilingen. Als er wel sprake lijkt te zijn van 'echte' verschillen, kan het optreden van non-respons alsnog roet in het eten gooien.