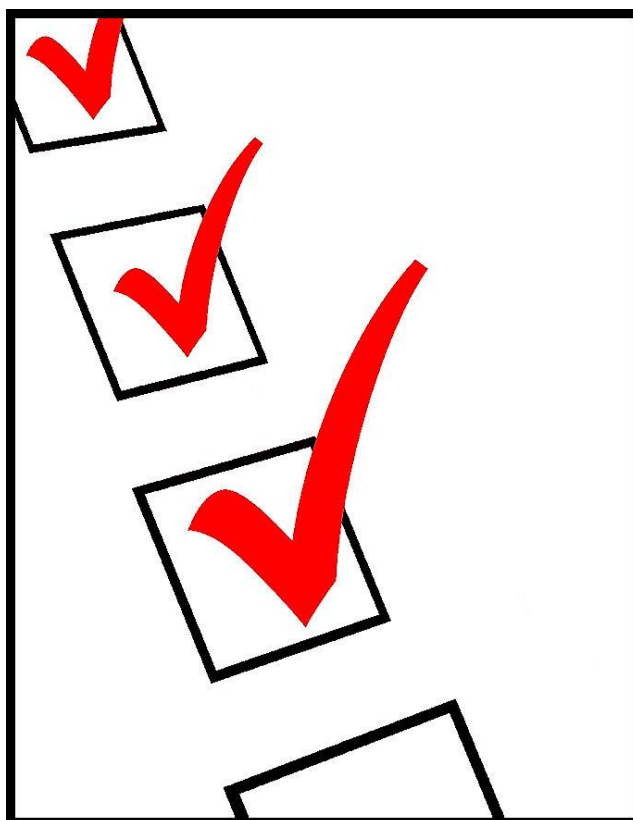


Checklist voor peilingen



Jelke Bethlehem

Versie 3.0 (april 2021)

Checklist voor peilingen

Onderzoeksrapport	Is er een <i>onderzoeksrapport</i> waarin staat uitgelegd hoe de peiling is opgezet en uitgevoerd? Zo niet, dan kun je niet vaststellen hoe goed of slecht de peiling is. Je kunt de peiling dan maar beter negeren.
Opdrachtgever en/of uitvoerder	Is het duidelijk wie de <i>opdrachtgever</i> en/of <i>uitvoerder</i> van de peiling is? Hebben die belang bij bepaalde uitkomsten? Dit zou kunnen leiden tot sturende vragen. Controleer de vragenlijst!
Doelpopulatie	Is de <i>doelpopulatie</i> van de peiling duidelijk omschreven? Dit is de groep mensen waaruit de steekproef is getrokken en waarop de uitkomsten van de peiling betrekking hebben.
Vragenlijst	<ul style="list-style-type: none"> • Is de <i>vragenlijst</i> opgenomen in het onderzoeksrapport? • Is de vragenlijst getest voordat hij het veld in ging? • Zijn de vragen goed geformuleerd? • Kan de volgorde van de vragen invloed hebben op de uitkomsten?
Steekproef	<ul style="list-style-type: none"> • Is de <i>steekproef</i> een aselecte steekproef uit de doelpopulatie waarin alle personen dezelfde kans op selectie hebben? • Is de steekproef een aselecte steekproef uit een deel van de doelpopulatie (bijvoorbeeld alleen de leden van een beroepsvereniging)? Dan hebben de uitkomsten alleen betrekking op die deelpopulatie. Er is dan sprake van <i>onderdekking</i> (<i>under-coverage</i>). • Is er sprake van een quotasteekproef. Die is vaak niet representatief. • Is er sprake van een online steekproef met zelfselectie? Die is meestal ook niet representatief. • Is er sprake van een steekproef uit een online panel? Dan is die steekproef alleen representatief als het panel ook representatief is.
Respons	Is vermeld hoeveel personen zijn benaderd voor het onderzoek? En hoeveel daarvan uiteindelijk meededen (de <i>respons</i>). Als je niet weet hoe groot de respons is, kun je ook niet bepalen hoe precies de schattingen zijn.
Non-respons	Is het <i>percentage non-respons</i> vermeld? Is dit percentage voldoende laag, zeg lager dan 50%? Hoge non-respons kan leiden tot verkeerde conclusies.
Wegen	<ul style="list-style-type: none"> • Is de peiling met een weging gecorrigeerd voor een mogelijk gebrek aan representativiteit? • Welke variabelen zijn gebruikt voor deze <i>weging</i>. Is die weging voldoende om vertekening van de uitkomsten te verkleinen?
Onzekerheidsmarges	<ul style="list-style-type: none"> • Is gemeld hoe groot de <i>onzekerheidsmarges</i> van de uitkomsten zijn? • Is aangegeven dat afwijkingen nog groter kunnen zijn als gevolg van non-respons.

Inleiding

Er zijn veel peilingen in Nederland. Dat merk je vooral 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 vragen peilingen steeds vaker de mening van ‘de Nederlander’ over allerlei onderwerpen.

Peilingen, enquêtes en surveys: het zijn verschillende namen voor hetzelfde soort onderzoek. Aan een *steekproef* van personen uit een doelgroep leg je een *vragenlijst* voor. Die vragen kunnen gaan over feitelijke zaken, gedrag en meningen. Met de verkregen antwoorden probeer je dan uitspraken te doen over de doelgroep als geheel. Dat kan, maar dan moet je het onderzoek wel op wetenschappelijk verantwoorde wijze hebben opgezet en uitgevoerd. Met deze checklist voor peilingen kun je vaststellen of dit inderdaad het geval is.

Voorals het internet is er de oorzaak van dat het aantal peilingen sterk is toegenomen. Internet maakt het mogelijk om eenvoudig, snel en goedkoop bij heel veel mensen gegevens te verzamelen. De vraag is echter of al die peilingen wel een goed beeld geven van de werkelijkheid. Vormen de uitkomsten van een peiling een goede afspiegeling van de doelgroep? Anders gezegd: zijn die peilingen wel *representatief*?

Er zijn dus steeds meer peilingen. Daar zitten goede en slechte peilingen bij. Het is niet eenvoudig voor gebruikers van de uitkomsten van peilingen (onderzoekers, journalisten, bestuurders, beleidsmakers, politici en ook het algemene publiek) om op simpele wijze het kaf van het koren te scheiden. Deze checklist helpt je daarbij. Door de vragen in de checklist één voor één af te lopen en te beantwoorden, krijg je een goede eerste indruk van de kwaliteit van de peiling. 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 het onderzoek en de uitkomsten, dan is het verstandig er verder geen aandacht aan te besteden. Of je kunt bij de beschrijving van het onderzoek waarschuwen voor mogelijke problemen.

Er is geprobeerd de checklist simpel te houden. Daardoor heeft hij een vrij uitgesproken karakter gekregen: een aspect van de peiling is goed of fout en er is geen tussenweg. Uiteraard is de werkelijkheid vaak ingewikkelder. Dan zal het eindoordeel ook wat genuanceerder zijn.

Hieronder lichten we de vragen in de checklist nader toe. Uit die uitleg zal blijken waarom beantwoording van de vragen belangrijk is.

Dit is de derde versie van de checklist voor peilingen. De eerste versie kwam uit in 2012. Die checklist was een gezamenlijk initiatief van het *Nederlandstalig Platform voor Survey-onderzoek* (NPSO), het *Centraal Bureau voor de Statistiek* (CBS) en de *Vereniging voor Onderzoeksjournalisten* (VVOJ).

Vraag 1: is er een onderzoeksrapport beschikbaar?

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

Het onderzoeksrapport moet minimaal de volgende aspecten beschrijven:

- Wie is de *opdrachtgever* van de peiling?
- Welke *onderzoeksbureau* heeft de peiling uitgevoerd?
- Wat is de *doelpopulatie*? Dat is de doelgroep waaruit de steekproef is getrokken en waarop de uitkomsten van het onderzoek betrekking hebben.
- De volledige *vragenlijst*. Het moet duidelijk zijn of de vragenlijst van tevoren is getest.
- Het gebruikte *steekproefkader*. Dit is de lijst waarin voor ieder lid van de doelpopulatie de contactgegevens staan. Denk bijvoorbeeld aan naam, adres, telefoonnummer of e-mailadres. Het is de lijst waaruit je de steekproef trekt. Problemen kunnen ontstaan als er sprake is van

onderdekking / under-coverage (niet alle leden van de doelpopulatie zitten in het steekproefkader) of *overdekking / over-coverage* (in het steekproefkader zitten personen die niet thuis horen in de doelpopulatie).

- De wijze waarop de *steekproef* is getrokken. Het moet duidelijk zijn of de steekproef is geloot en hoe de steekproef 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* ($100 \times \text{Omvang respons} / \text{Omvang bruto steekproef}$).
- De wijze waarop de respons is *gecorrigeerd* voor *non-respons* (en eventuele andere selectie-effecten). Welke variabelen (*weegvariabelen*) zijn gebruikt voor de weging? En hoe zijn die weegvariabelen gebruikt bij het berekenen van de gewichten (het *weegmodel*)?
- De omvang van de *onzekerheidsmarges*. Merk op dat je deze alleen kunt berekenen 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. Is dit gemeld in het rapport?

Vraag 2: 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 zie je regelmatig persberichten 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. Het is een geval van “Wij van WC-Eend adviseren WC-Eend”.

De Britse omroep BBC heeft 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 het bericht aangeven wie de opdrachtgever is. Uiteraard moet dan wel bekend zijn wie de opdrachtgever is.

Vraag 3: wat is de doelpopulatie?

De *doelpopulatie* is de groep van mensen waarop de uitkomsten van het onderzoek betrekking hebben. De doelpopulatie moet duidelijk zijn omschreven.

Je trekt de steekproef uit een *steekproefkader*. Dit is een lijst waarin voor ieder lid van de doelpopulatie de contactgegevens staan. Denk bijvoorbeeld aan naam, adres, telefoonnummer en e-mailadres. Het is de lijst waaruit je de steekproef trekt.

Problemen kunnen ontstaan als het steekproefkader niet de hele doelpopulatie afdekt. In feite trek je dan de steekproef dan uit een andere, kleinere groep. We noemen dat *onderdekking (under-coverage)*. De consequentie daarvan is dat de uitkomsten alleen betrekking hebben op die kleinere groep en niet op de oorspronkelijke doelpopulatie. Je kunt niet generaliseren naar de doelpopulatie.

Voorbeeld: een patiëntenvereniging wil een peiling doen onder patiënten die lijden aan een bepaalde ziekte. Daarvoor trekt de vereniging een steekproef uit haar leden. Als gevolg daarvan bestaat de doelpopulatie niet meer uit alle patiënten die lijden aan die ziekte, maar alleen uit de patiënten die lid zijn van de vereniging. Die hoeven niet representatief te zijn voor alle patiënten.

Vraag 4: hoe is 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 zijn vermeden:

- Onbegrijpelijke vragen. Vragen kunnen onbegrijpelijk zijn voor de respondenten. De oorzaak kan het gebruikte jargon zijn, onbekend woordgebruik of lange, vage of ingewikkelde zinsconstructies. Voorbeeld: “*Bent u tevreden over de recreatieve voorzieningen in uw woonplaats?*”.
- Vragen die je op verschillende manieren kunt interpreteren. Voorbeeld: “*Wanneer ging u van school af?*”. Gaat het hier om een datum, een leeftijd of andere omstandigheden (toen ik in verwachting was)?
- Suggestieve en sturende vragen. Voorbeeld: “*Vindt u ook niet dat ... ?*”.
- Dubbele vragen. Dat zijn vragen die meer dan één vraag stellen. 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 je vraagt naar gebeurtenissen die niet heel belangrijk zijn en die regelmatig voorkomen. Voorbeeld: “*Hoe vaak heeft u in het afgelopen jaar een bezoek gebracht aan een supermarkt?*”.

Voor een goede peiling is het noodzakelijk dat je de vragenlijst eerst test voordat je hem echt gebruikt in een peiling. Dus, is de vragenlijst getest?

Vraag 5: hoe is de steekproef getrokken?

Om op verantwoorde wijze conclusies te kunnen trekken over de doelpopulatie, moet je de steekproef trekken door *loting*. Iedereen in de populatie moet een positieve kans hebben gehad om in de steekproef te komen. Deze trekkingskansen moeten bekend zijn (of je moet ze kunnen berekenen).

De eenvoudigste wijze van trekken is het loten van een *aselecte steekproef*. Daarbij heeft iedereen in de doelpopulatie *dezelfde kans* om in de steekproef te komen. Dit leidt tot een *representatieve steekproef*. Daarvoor geldt het principe dat het percentage (of het gemiddelde) in de steekproef een goede schatter is voor het percentage (of het gemiddelde) in de populatie. Het zijn valide schattingen. Ook kun je onzekerheidsmarges uitrekenen. Die geven aan hoeveel je schattingen op basis van de steekproef maximaal kunnen afwijken van de werkelijke cijfers.

Het is denkbaar om de steekproef met *ongelijke kansen* te trekken. Het berekenen van schattingen voor de doelpopulatie is dan wat ingewikkelder, omdat je moet corrigeren voor die ongelijke kansen. Een voorbeeld is een steekproef van personen die je krijgt 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.

Je trekt de steekproef uit een *steekproefkader*. Dit is de (elektronische) lijst waarin voor ieder lid van de doelpopulatie de contactgegevens staan (bijvoorbeeld naam, adres, telefoonnummer of e-mailadres). Als het steekproefkader niet de gehele doelpopulatie bestrijkt (een geval van *onderdekking*), dan hebben de conclusies van het onderzoek alleen betrekking op het deel van de doelpopulatie dat je via het steekproefkader bereikt.

Als de steekproef niet is verkregen door loting, dan zijn de trekkingskansen onbekend en kun je nooit goede schattingen voor populatiepercentages (en populatiegemiddelden) berekenen. Dit is bijvoorbeeld het geval bij online peilingen waarbij de steekproef is geworven door *zelfselectie*. Daarbij staat de vragenlijst op het internet en de respondenten zijn mensen die de peiling toevallig tegenkomen, het leuk vinden om mee te doen aan een peiling, geïnteresseerd zijn in het onderwerp van de peiling en spontaan besluiten de vragenlijst in te vullen. Met zelfselectie krijg je daarom geen representatieve steekproef. De uitkomsten zijn niet valide. En je kunt ook geen onzekerheidsmarges uitrekenen.

Bij steekproeven op basis van zelfselectie kunnen zich nog meer problemen voordoen. Zo is het soms mogelijk voor de respondenten om de vragenlijst meer dan eens in te vullen. Ook kunnen personen van buiten de doelpopulatie soms meedoen aan de peiling. En er zijn gevallen bekend waarbij een groep deelnemers bewust probeert de uitkomsten te manipuleren. Dit soort problemen kunnen zich niet voordoen bij aselechte steekproeven.

Diverse marktonderzoekbureaus maken voor hun peilingen gebruik van een *online panel*. Een dergelijk panel bestaat uit een grote groep personen die hebben aangegeven regelmatig te willen meedoen aan peilingen (al of niet tegen betaling). Aangezien van elk lid van het panel het e-mailadres beschikbaar is, kun je snel, eenvoudig en goedkoop een steekproef trekken uit een online panel. Er zijn twee soorten online panels: panels die gebaseerd zijn op een aselechte steekproef uit de doelpopulatie en panels die gevuld zijn op basis van zelfselectie. Het moet uit het onderzoeksrapport duidelijk zijn hoe het gebruikte online panel is samengesteld. Bedenk dat een aselechte steekproef uit een online panel alleen maar representatief is als het gebruikte online panel representatief is.

Vraag 6: hoe groot is de steekproef?

Als de steekproef door loting is verkregen, dan kun je de *precisie* van de schattingen berekenen. Daarbij geldt de eenvoudige regel dat de precisie toeneemt als de steekproef groter is. Daarom is het van belang de omvang van de steekproef te vermelden.

Je geeft de precisie meestal weer 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 onzekerheidsmarges.

Tabel 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

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

Voorbeeld: 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 zijn door de 'ruis' in de steekproef. Je kunt dus niet concluderen dat er sprake is van een significante toename van het percentage stemmers.

Vraag 7: hoe groot is de non-respons?

Non-respons is het verschijnsel dat in de steekproef getrokken personen niet de gewenste gegevens verstrekken. Dat kan verschillende oorzaken hebben: het lukt niet contact met ze te leggen

(bijvoorbeeld omdat ze zelden thuis zijn), ze weigeren mee te doen aan de peiling of ze zijn niet in staat om de vragen te beantwoorden (door bijvoorbeeld ziekte of taalproblemen). Non-respons leidt er heel vaak toe dat bepaalde groepen mensen over- of ondervertegenwoordigd zijn in de steekproef. Daardoor is de gerealiseerde steekproef geen goede afspiegeling meer van de doelpopulatie. De steekproef is dus niet meer representatief. Drie factoren bepalen de omvang van de vertekening in de uitkomsten:

- Hoeveel mensen doen niet mee? Een groter percentage non-respons leidt tot een grotere afwijking in de uitkomsten. Is de respons bij een peiling dus laag, dan is het risico groot dat er ernstige afwijkingen zitten 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 daarvan schat je de opkomst te hoog in.
- Doen sommige groepen mensen vaker aan een peiling mee dan andere groepen? 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 je de antwoorden van de non-respondenten op de vragen kent. Omdat het echter om non-respondenten gaat, zijn die antwoorden onbekend. Het is wel mogelijk om uit te rekenen hoe groot de afwijking ten gevolge van non-respons maximaal kan zijn (de *maximale vertekening*).

Stel eens dat slechts 40% van de personen in de steekproef respondeert in een verkiezingsonderzoek. Van die respondenten zegt 60% te 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 juist 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. De schatting voor het werkelijke percentage stemmers zal ergens tussen de 24% en 84% liggen. Een dergelijke grote bandbreedte voor het steekproefpercentage maakt dit onderzoek in feite onbruikbaar. Gelukkig doen die extreme situaties zich in de praktijk niet vaak voor, maar het effect van non-respons kan toch groot zijn.

Tabel 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

Tabel 2 bevat de bandbreedte voor de steekproefpercentages 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.

Vraag 8: is er gecorrigeerd voor non-respons?

Het is belangrijk om te corrigeren voor de effecten van non-respons door het uitvoeren van een *weging*: aan elke respondent ken je een gewicht toe. Die gewichten zijn zo berekend dat ze corrigeren voor de over- of ondervertegenwoordiging van groepen in de respons.

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 gerepsondeerd in de peiling want ze zijn oververtegenwoordigd. Om hiervoor te corrigeren, krijgt elke responderende man een gewicht van

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

Dit is dus het populatiepercentage gedeeld door het steekproefpercentage. Elke man telt nog maar meetelt voor 0,825 in plaats van 1. Het gewicht is kleiner dan 1 omdat er teveel mannen zitten 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 zitten in de steekproef.

Het toekennen van gewichten aan mannen en vrouwen maakt de respons representatief met betrekking tot het geslacht. Die gewichten kun je uitrekenen omdat de juiste percentages in de doelpopulatie bekend zijn. Het idee achter wegen is nu om de respons representatief te maken met betrekking tot zoveel mogelijk variabelen. Noodgedwongen kunnen dat alleen variabelen zijn die in de peiling zijn gemeten en waarvoor de verdeling in de populatie bekend is. Voorbeelden van veel gebruikte variabelen zijn geslacht, leeftijd, burgerlijke staat, opleidingsniveau, samenstelling van het huishouden, regio en stemgedrag bij de vorige verkiezingen.

De hoop is dat als je de respons representatief maakt met betrekking tot zoveel mogelijk variabelen, hij misschien ook wel representatief is met betrekking tot de verschijnselen die je onderzoekt in de peiling. Daarmee zou dan de vertekening in de uitkomsten verdwijnen.

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

- Er is een sterke samenhang tussen het verschijnsel dat je onderzoekt (de onderzoeksvariabelen) en de variabelen die je gebruikt voor wegen (de weegvariabelen).
- Er is een sterke samenhang tussen het responsgedrag en de variabelen die je gebruikt voor wegen (de weegvariabelen).

Soms heeft wegen veel effect, bijvoorbeeld als je bij een onderzoek naar tijdbesteding weegt 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 naar leeftijd en een opiniepeiling is vaak zinloos, omdat die variabelen de opinies niet of nauwelijks bepalen. In alle gevallen is het van belang om na te gaan welke groepen onder- of oververtegenwoordigd zijn, wat dat voor effect dat op de uitkomsten kan hebben en of wegen de uitkomsten van het onderzoek beïnvloed.

Vraag 9: zijn er onzekerheidsmarges gegeven?

De uitkomsten van een peiling zijn slechts schattingen van wat er in de doelpopulatie aan de hand is. Het is daarom onrealistisch te veronderstellen dat die uitkomsten exact overeenkomen met de werkelijk cijfers voor de hele doelpopulatie. Zelfs in het ideale geval van een nette aselechte steekproef 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 kun je tabel 1 gebruiken.

Als de onderzoekers geen onzekerheidsmarges vermelden, kun je ze soms toch nog met de hand uitrekenen. Als je het percentage in de steekproef aangeeft met p en de omvang van de steekproef met n , dan is de onzekerheidsmarge (bij benadering) gelijk aan

$$M = 1,96 \times \sqrt{\frac{p \times (100 - p)}{n - 1}}.$$

Voorbeeld: voor een peiling trek je een aselechte steekproef van 500 personen. Iedereen doet mee. Er is geen non-respons. Van de respondenten zegt 60% voorstander te zijn van een bepaalde maatregel. De bijbehorende onzekerheidsmarge is dan gelijk aan

$$M = 1,96 \times \sqrt{\frac{60 \times (100 - 60)}{499}} = 4,3.$$

Het percentage voorstanders in de doelpopulatie zal met hoge waarschijnlijkheid liggen tussen de $60 - 4,3 = 55,7\%$ en de $60 + 4,3 = 64,3\%$.

Omdat je de steekproef loot, kunnen de uitkomsten dus afwijken van de werkelijke cijfers in de populatie. De maximale afwijking kun je uitrekenen. Dat is de onzekerheidsmarge. Lastiger is het dat het optreden van non-respons nog tot een extra afwijking kan leiden. Over die afwijking kun je echter niets zeggen. Je moet dus in ieder geval de onzekerheidsmarges van de uitkomsten melden. En daar moet je aan toevoegen dat de afwijking nog groter kan zijn als gevolg van het optreden van non-respons. En die extra afwijking is groter naarmate het percentage non-respons groter is.

Voorbeeld 1: Social Media Stress

Op maandag 7 mei 2012 melden verschillende media (waaronder NOS, Trouw, De Telegraaf en Metro) dat jongeren tussen 13 en 18 jaar lijden aan een serieuze vorm van *Social Media Stress* (SMS). De sociale media blijken met hun subtiele stimuli zoals geluiden, pushberichten, aandacht en beloningen jongeren in hun greep te houden. Dat zou blijken uit een peiling van de Nationale Academie voor Media & Maatschappij.

Volgens deze peiling geven jongeren aan niet meer zelfstandig te kunnen stoppen, omdat zij bang zijn buitengesloten te raken. Wanneer deze angst ernstige vormen aanneemt, kun je spreken van *Fear Of Missing Out* (FOMO). De Nationale Academie voor Media & Maatschappij pleit voor een nieuw bewustzijn en betere begeleiding van jongeren.

Dit lijkt een ernstige zaak, maar kloppen de beweringen ook? Om deze vraag te beantwoorden, gebruiken we de checklist. We lopen de vragen één voor één af.

Vraag 1: is er een onderzoeksrapport beschikbaar?

Op de website van de Academie staat een onderzoeksrapport. De beschrijving van de opzet en de uitvoering van de peiling is echter zeer summier. De informatie is onvoldoende om te kunnen vaststellen of het om wetenschappelijk verantwoorde peiling gaat. Na een verzoek daartoe per e-mail sturen de onderzoekers meer informatie. Daaruit blijkt dat deze peiling diverse tekortkomingen heeft.

Vraag 2: wie is de opdrachtgever?

De opdrachtgever en uitvoerder van het onderzoek is de *Nationale Academie voor Media en Maatschappij*. De twee directeuren van deze stichting zijn ook de opstellers van het onderzoeksrapport. Ze zijn ook de hoofddocenten van deze Academie. Die academie geeft cursussen op het terrein van de sociale media, zoals bijvoorbeeld een cursus 'Social Media Professional'. De opdrachtgever van het onderzoek heeft dus belang bij de uitkomsten ervan: naarmate de problemen met sociale media groter zijn, is er meer behoefte aan cursussen op dit terrein.

Vraag 3: wat is de doelpopulatie?

In eerste instantie lijkt het onderzoek betrekking te hebben op alle jongeren tussen de 13 en 18 jaar. Het onderzoeksrapport beperkt zich echter tot jongeren in die leeftijdsgroep die de beschikking hebben over een smartphone met internetverbinding. Volgens het rapport zou dat 76% van de jongeren zijn. Het blijft onduidelijk wat nu precies de doelpopulatie is.

Vraag 4: hoe is de kwaliteit van de vragenlijst?

De vragenlijst is niet opgenomen in het onderzoeksrapport. Dus daarover valt weinig te zeggen. Als je de analyse van de uitkomsten leest, dan krijg je wel het gevoel dat alles erop gericht is om aan te tonen dat jongeren leiden aan Social Media Stress.

Vraag 5: hoe is de steekproef getrokken?

Uit nadere informatie van de Academie blijkt dat de steekproef niet netjes via loting is getrokken. De Academie beschikt volgens eigen zeggen over een groot netwerk van meer dan 850 jeugdprofessionals die, over het hele land verspreid, werken in het onderwijs, bibliotheken en mediatheken, jeugdhulpverlening en buurtwerk. Aan een selectie van deze jeugdprofessionals is gevraagd in hun eigen werkomgeving willekeurig jongeren in de leeftijd van 13 tot en met 17 jaar uit te nodigen om aan het onderzoek mee te werken. Er wordt vooraf niet verteld waar het onderzoek over gaat. De jongeren vullen de vragenlijst online in. Er doen zo'n 25 instanties mee met ongeveer 20 jongeren per instantie. Dit is een merkwaardige manier van trekken van de steekproef. Je kunt je afvragen voor welke populatie de steekproef representatief zou moeten zijn. Het netwerk van jongeren rondom jeugdprofessionals? En dan nog, er is geen aselechte steekproef getrokken.

Vraag 6: hoe groot is de steekproef?

De steekproef bestaat uiteindelijk uit 493 jongeren (240 jongens en 253 meisjes). Dat is dus een vrij kleine steekproef. Je moet daarom rekening houden met ruime onzekerheidsmarges (zelfs als de steekproef netjes is geloot).

Vraag 7: hoe groot is de non-respons?

Het hele probleem van de non-respons is genegeerd. Wat is er gedaan met de jongeren die wel werden uitgenodigd om mee te doen aan het onderzoek, maar dat niet wilden? Is er sprake van selectieve non-respons? Zijn die jongeren anders? Hebben ze, bijvoorbeeld, misschien geen smartphone? De onderzoekers hebben de non-respons niet in beeld gebracht er al helemaal niet voor gecorrigeerd.

Vraag 8: is er gecorrigeerd voor non-respons?

De onderzoekers hebben de uitkomsten gewogen. Sterker nog, ze zeggen dat ze een 'CBS-weging' hebben uitgevoerd en dat daarom het onderzoek representatief is. Bij navraag blijkt echter dat er alleen is gewogen naar geslacht. Dat is niet nodig, want er zitten al ongeveer evenveel jongens en meisjes in de steekproef. Dus de weging doet helemaal niets.

Je komt wel vaker onderzoekers tegen die claimen dat hun uitkomsten goed zijn omdat er met CBS-cijfers is gewogen. Het CBS levert hiervoor zelfs een hele reeks tabellen aan de MOA, de brancheorganisatie van marktonderzoekers. Die wordt de 'Gouden Standaard' genoemd. Je kunt je afvragen of het CBS in dit geval wel blij moet zijn met dergelijke claims.

Vraag 9: zijn er onzekerheidsmarges gegeven?

De onderzoekers van de Nationale Academie voor Media & Maatschappij vermelden geen onzekerheidsmarges bij hun uitkomsten. Dat kan ook niet omdat ze geen aselechte steekproef hebben

getrokken. Maar als eerste indicatie zou je toch minsten de onzekerheidsmarges kunnen vermelden die uitgaan van een aselechte steekproef.

Overzicht

Tabel 3 hieronder vat de antwoorden op de vragen van de checklist samen. Voor acht van de negen vragen is de conclusie: Niet OK. We kunnen dus concluderen dat dit geen goede peiling is.

Tabel 3. Checklist voor de peiling naar Social Media Stress

1.	Onderzoeksrapport?	Is er is geen onderzoeksrapport.	Niet OK
2	Opdrachtgever en/of uitvoerder?	Nationale Academie voor Media & Maatschappij heeft belang bij bepaalde uitkomsten.	Niet OK
3.	Doelpopulatie?	De doelpopulatie is onduidelijk. Mogelijk alle jongeren tussen de 13 en 18 jaar met een smartphone.	Niet OK
4.	Vragenlijst?	De vragenlijst is niet opgenomen in het onderzoeksrapport. Mogelijk bevat hij sturende vragen.	Niet OK
5.	Steekproef?	Er is geen aselechte steekproef getrokken.	Niet OK
6.	Respons?	Omvang van steekproef is 493.	OK
7.	Non-respons?	Het percentage non-respons is niet gemeld.	Niet OK
8.	Weging?	Er is gewogen naar geslacht. Deze weging is onvoldoende.	Niet OK
9.	Onzekerheidsmarges?	Er zijn geen onzekerheidsmarges gemeld.	Niet OK

Voorbeeld 2: thuisblijven met oud en nieuw

Stand.nl is een radioprogramma dat elke werkdag wordt uitgezonden op de zender NPO Radio 1. In het programma komt dagelijks een actuele kwestie aan de orde met een actuele stelling. De luisteraars kunnen over de stelling online of per telefoon stemmen (*eens* of *oneens*). Als peiling zit *Stand.nl* niet goed in elkaar. Het is geen representatieve peiling en dus kun je de uitkomsten niet generaliseren naar de hele bevolking. En ook niet naar alle luisteraars. We illustreren de tekortkomingen van deze peilingen aan de hand van een voorbeeld van een peiling van *Stand.nl*.

Op 30 december 2020 is de stelling “Iedereen moet uit solidariteit thuisblijven met oud en nieuw”. De politie heeft aangegeven hard te zullen optreden, zeker met het oog op het zware, illegale vuurwerk dat zou kunnen worden afgestoken. Ook minister Grapperhaus laat weten dat de politie ‘lik-op-stuk’ geeft tijdens de jaarwisseling. Het is sowieso de bedoeling dat mensen zo veel mogelijk thuis blijven, zeker met het oog op de zorg die in die tijden van Corona tegen zijn grenzen aanloopt.

Nadat bijna 1300 respondenten gereageerd hebben op de stelling, blijkt dat 56% voor de stelling is. De vraag is nu of je deze uitkomst kunt generaliseren naar de hele bevolking. Om dit te kunnen vaststellen, lopen we door de checklist heen.

Vraag 1: is er een onderzoeksrapport beschikbaar?

Er is geen onderzoeksrapport beschikbaar. Uit summiere aanwijzingen op de internetpagina van het programma *Stand.nl* kun je een globaal beeld krijgen van hoe de peiling is uitgevoerd.

Vraag 2: wie is de opdrachtgever?

De peiling is opgezet en uitgevoerd door de redactie van het programma Stand.nl. Die redactie is dus de opdrachtgever van het onderzoek. Het lijkt niet waarschijnlijk dat deze redactie belang heeft bij bepaalde uitkomsten van het onderzoek. Het gevaar van sturende vragen is dus niet groot.

Vraag 3: wat is de doelpopulatie?

Het is helemaal niet duidelijk wat de doelpopulatie van de peiling is. Er zijn diverse mogelijkheden:

- Alle inwoners van Nederland.
- Alle inwoners van Nederland vanaf een bepaalde leeftijd, bijvoorbeeld vanaf 18 jaar.
- Alle luisteraars naar de radiozender NPO Radio 1.
- Alle luisteraars naar het radioprogramma Stand.nl.

Merk op dat de vragenlijst op het internet staat en dat iedereen die kan invullen, wat de doelpopulatie van de peiling ook is. Omdat iedereen kan meedoen aan de peiling, gaan we er maar vanuit dat de doelpopulatie bestaat uit alle inwoners.

Vraag 4: hoe is de kwaliteit van de vragenlijst?

De vragenlijst van de peiling bevat slechts drie vragen: de mening over de stelling, geslacht en leeftijd in vier categorieën (0-20 jaar, 20-40 jaar, 40-60 jaar en 60 jaar en ouder). Er is een probleem met die vier categorieën. Als je 20 jaar bent, hoor je dan in eerste categorie (0-20 jaar) of in de tweede categorie (20-40 jaar)? Hetzelfde probleem doet zich voor bij respondenten van 40 en 60 jaar.

Vermoedelijk gaat het bij 0-20 jaar om personen met een leeftijd van 0 tot 20 jaar. Dus de 20-jarigen horen er niet bij. Dit zitten in de categorie 20-40 jaar. Het is beter om in de aanduiding van de categorieën niet een streepje te gebruiken maar het woordje 'tot': 0 tot 20 jaar.

Vraag 5: hoe is de steekproef getrokken?

Bij voorkeur trek je als onderzoeker een aselechte steekproef. Daarbij heeft iedereen dezelfde kans om in de steekproef te komen. Je steekproef is dan dus representatief. Bij Stand.nl wordt gebruikt gemaakt van zelfselectie. Iedereen die dat wil, kan meedoen. Je krijgt dan alleen mensen in de steekproef die spontaan besluiten mee te doen. Die dus gemotiveerd en geïnteresseerd zijn. Het zijn vooral mensen die elke ochtend naar het radioprogramma luisteren. Dat leidt niet tot een representatieve steekproef.

Vraag 6: hoe groot is de steekproef?

Je kunt twee uur lang de vragenlijst van de peiling invoeren, van 9.30 uur tot 11.30 uur. Op elk willekeurig moment kun je de tussenstand van de uitkomsten bekijken. Wij hebben de uitkomsten geanalyseerd nadat bijna 1300 respondenten gereageerd hebben op de stelling. Op dat moment blijkt 56% voor de stelling te zijn.

Vraag 7: hoe groot is de non-respons?

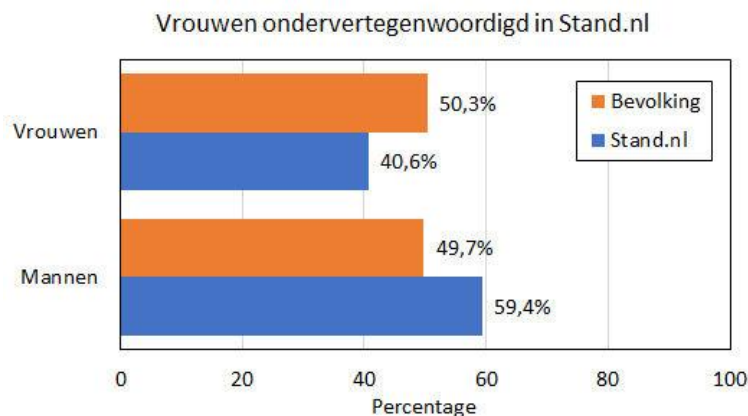
Het hele probleem van de non-respons is genegeerd. In principe zijn alle luisteraars naar het programma uitgenodigd om mee te doen aan de peiling. Slechts 1300 luisteraars hebben dat gedaan. De rest is dus non-respons. De vraag is nu of er sprake is van een selectieve respons. Zijn de respondenten anders dan de overige luisteraars? De redactie heeft de non-respons niet in beeld gebracht er al helemaal niet voor gecorrigeerd.

Vraag 8: is er gecorrigeerd voor non-respons?

Je kunt de variabelen geslacht en leeftijd gebruiken om na te gaan of er sprake is van selectieve respons. Je kunt vaststellen hoeveel mannen en vrouwen er in de peiling zitten en of de percentages mannen en vrouwen in de peiling overeen komen met de percentages mannen en vrouwen in de hele bevolking.

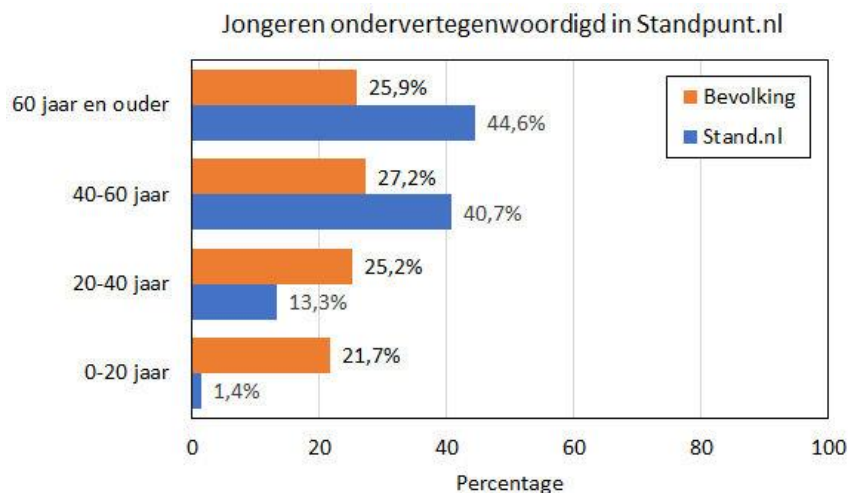
In grafiek 1 hieronder is die vergelijking gemaakt. De oranje staven geven de percentages mannen en vrouwen in de bevolking weer. Duidelijk is te zien dat er ongeveer evenveel mannen als vrouwen zijn. De blauwe staven staan voor de percentages mannen en vrouwen in de peiling. Er is duidelijk iets mis. Er zitten te weinig vrouwen in de peiling (40,6% in plaats van 50,3%) en er zitten teveel mannen in (59,4% in plaats van 49,7%). In beide gevallen is er sprake van een afwijking van ongeveer 10 procentpunten.

Grafiek 1. De verdeling van geslacht in de bevolking en in de peiling



We kunnen ook voor de variabele leeftijd de verdeling in de bevolking vergelijken met die in de peiling. Het resultaat staat in de grafiek 2 hieronder. De problemen met de representativiteit blijken ook hier groot te zijn.

Grafiek 2. De verdeling van leeftijd in de bevolking en in de peiling



De ouderen zijn zwaar oververtegenwoordigd in Standpunt.nl. In de peiling is 44,6% 60 jaar of ouder en in de bevolking is dit maar 25,9%. Een afwijking van 18,7 procentpunten dus! Ook de middelbaren tussen de 40 en 60 jaar zijn behoorlijk oververtegenwoordigd. Er is hier sprake van een afwijking van 12,8 procentpunten.

Bij de jongeren zie je het omgekeerde effect. Die jongeren (0-20 jaar en 20-40 jaar) zijn juist ondervertegenwoordigd in de peiling. De situatie is vooral ernstig voor de jongeren in de groep 0-20

jaar. De peiling zou voor 21,7% (één op de vijf) uit deze jongeren moeten bestaan, maar het is slechts een minimaal percentage van 1,4%.

Mannen zijn dus oververtegenwoordigd in de peiling en vrouwen zijn ondervertegenwoordigd. Dat betekent dat de mening van vrouwen onvoldoende doorklinkt in de resultaten. Ook zijn ouderen oververtegenwoordigd en jongeren ondervertegenwoordigd. De mening van jongeren klinkt dus onvoldoende door. We moeten concluderen dat deze peiling niet representatief is en we dus de uitkomsten niet kunnen generaliseren naar de hele bevolking.

Als een peiling niet representatief is, kun je proberen een correctie uit te voeren. Dat kan via een procedure die we *wegen* noemen. De mensen van Stand.nl hebben dat helaas niet gedaan. Het zou echter wel kunnen. We schetsen hier hoe je dit zou kunnen doen.

Omdat we verdeling naar geslacht en de verdeling naar leeftijd in de peiling en in de bevolking kennen, kunnen we gewichten bepalen die corrigeren voor onder- of over vertegenwoordiging van bepaalde groepen. Tabel 4 toont de resulterende gewichten.

Tabel 4. Gewichten

Geslacht	Leeftijd	Gewicht
Man	0 – 20 jaar	12,020
Man	20 – 40 jaar	1,670
Man	40 – 60 jaar	0,604
Man	60 jaar en ouder	0,433
Vrouw	0 – 20 jaar	22,913
Vrouw	20 – 40 jaar	2,184
Vrouw	40 – 60 jaar	0,749
Vrouw	60 jaar en ouder	0,838

Je kunt bijvoorbeeld zien dat jonge vrouwen (0-20 jaar) een groot gewicht van 22,913 krijgen. Dus elke jonge vrouw telt mee voor bijna 23 jonge vrouwen. Oude mannen (60 jaar en ouder) krijgen juist een klein gewicht van 0,433. Dus oude mannen tellen maar mee voor minder dan een halve oude man. Zo wordt de zaak, althans voor wat betreft geslacht en leeftijd, weer rechtgetrokken.

Als je nu opnieuw het percentage voorstanders van thuisblijven met oud en nieuw uitrekent (maar dan gewogen), kom je uit op 39%. Dat is een stuk lager dan het ongewogen percentage van 56%! We kunnen verwachten dat 39% een betere benadering is van het percentage voorstanders in de bevolking. Het oorspronkelijke percentage is dus veel te rooskleurig. Er is geen meerderheid van 56% voorstanders maar een meerderheid van 61% tegenstanders.

Standpunt.nl vraagt alleen naar geslacht en leeftijd. Daarom kun je alleen voor deze twee variabelen het gebrek aan representativiteit analyseren en corrigeren. Je kunt echter verwachten dat nog meer variabelen een rol spelen. Je zou bijvoorbeeld kunnen denken aan variabelen als opleidingsniveau en stemgedrag bij de vorige verkiezingen. Je kunt die kenmerken echter alleen gebruiken als je ze meet in de peiling. Helaas is dit niet het geval.

Vraag 9: zijn er onzekerheidsmarges gegeven?

De onderzoekers van Stand.nl vermelden geen onzekerheidsmarges bij hun uitkomsten. Dat kan ook niet omdat ze geen aselechte steekproef hebben getrokken. Als je veronderstelt dat er sprake is van een aselechte steekproef van 1300 personen, dan heeft het gecorrigeerde percentage van 39% een onzekerheidsmarge van 2,7 procentpunten. Dit zou betekenen dat het percentage in de doelpopulatie met grote waarschijnlijkheid ligt tussen de 36,3% en 41,7%.

Overzicht

Tabel 5 hieronder vat de antwoorden op de vragen van de checklist samen. Voor zeven van de negen vragen is de conclusie: Niet OK. We kunnen dus vaststellen dat dit geen goede peiling is.

Tabel 5. Checklist voor de peiling naar thuisblijven met oud en nieuw

1.	Onderzoeksrapport?	Is er is geen onderzoeksrapport.	Niet OK
2	Opdrachtgever en/of uitvoerder?	De redactie van Stand.nl heeft geen belang bij bepaalde uitkomsten.	OK
3.	Doelpopulatie?	De doelpopulatie is onduidelijk. Misschien de hele bevolking? Of alleen alle luisteraars?	Niet OK
4.	Vragenlijst?	Er zijn slechts drie vragen: mening, geslacht en leeftijd. Er is een probleem met de indeling in categorieën van leeftijd	Niet OK
5.	Steekproef?	Er is geen aselechte steekproef getrokken. Het is een geval van zelfselectie.	Niet OK
6.	Respons?	Omvang van steekproef is ongeveer 1300 personen.	OK
7.	Non-respons?	Het percentage non-respons is niet gemeld.	Niet OK
8.	Weging?	Er is niet gewogen. Weging naar geslacht en leeftijd is wel mogelijk. Zou tot een andere uitkomst hebben geleid.	Niet OK
9.	Onzekerheidsmarges?	Er zijn geen onzekerheidsmarges bepaald. Dan kan ook niet vanwege zelfselectie.	Niet OK

Boeken over peilingen

Bethlehem, J.G. (2009), *Applied Survey Methods: A Statistical Perspective*. John Wiley & Sons, Hoboken, New Jersey.

Bethlehem, J.G. (2018), *Understanding Public Opinion Polls*. CRC Press / Taylor & Francis, Boca Raton, Florida.