



**Universiteit  
Leiden**  
The Netherlands

**Het meten van de gevolgen van aardbevingen op de leefbaarheid in Groningen: een methodologische review naar leefbaarheidsonderzoeken. Universiteit Utrecht/Universiteit Leiden onderzoeksrapport in opdracht van Nationaal Coördinator Groningen (NCG)**

Lugtig, P.; Mors, E. ter; Bethlehem, J.G.

**Citation**

Lugtig, P., Mors, E. ter, & Bethlehem, J. G. (2017). *Het meten van de gevolgen van aardbevingen op de leefbaarheid in Groningen: een methodologische review naar leefbaarheidsonderzoeken. Universiteit Utrecht/Universiteit Leiden onderzoeksrapport in opdracht van Nationaal Coördinator Groningen (NCG)*. Utrecht: Utrecht University. Retrieved from <https://hdl.handle.net/1887/3630071>

Version: Publisher's Version

License: [Leiden University Non-exclusive license](#)

Downloaded from: <https://hdl.handle.net/1887/3630071>

**Note:** To cite this publication please use the final published version (if applicable).



Universiteit  
Leiden



Universiteit Utrecht

# Eindrapport

## Het meten van de gevolgen van aardbevingen op de leefbaarheid in Groningen

Een methodologische review naar  
leefbaarheidsonderzoeken

*7 juli 2017*

Peter Lugtig  
Emma ter Mors  
Jelke Bethlehem

Universiteit Utrecht  
Universiteit Leiden  
Universiteit Leiden



## **Het meten van de gevolgen van aardbevingen op de leefbaarheid in Groningen**

Een methodologische review naar leefbaarheidsonderzoeken  
Universiteit Leiden en Universiteit Utrecht

Onderzoek uitgevoerd door de afdeling Sociale en Organisationspsychologie van de  
Universiteit Leiden en de afdeling Methodologie en Statistiek van de Universiteit  
Utrecht, in opdracht van de Nationaal Coördinator Groningen (NCG)

Dr. Peter Lugtig (Universiteit Utrecht)  
Dr. Emma ter Mors (Universiteit Leiden)  
Prof. dr. Jelke Bethlehem (Universiteit Leiden)

*7 juli 2017 [Eindrapport]*

**Met dank aan:**

Dit rapport is tot stand gekomen met input, ondersteuning en meedenken van medewerkers van de afdeling Sociale en Organisationspsychologie van de Universiteit Leiden en van de afdeling Methodologie en Statistiek van de Universiteit Utrecht. Met speciale dank aan Jildou Talman van de Universiteit Leiden voor hulp bij het verzamelen van achtergrondinformatie en het in elkaar zetten van dit rapport. In het bijzonder willen we de onderzoekers bedanken die hebben gewerkt aan de vijf onderzoeken die worden gereviewd. Als onderdeel van ons onderzoek hebben we gesproken met Tom Postmes en Katherine Stroebe (Rijksuniversiteit Groningen), Femke de Haan en Marja Janssens (onderzoeksbureau CMO STAMM), Peter Boelhouwer en Joris Hoekstra (OTB Technische Universiteit Delft) en Stefan van der Veen en Dionne Faber (onderzoeksbureau APE Public Economics). Zij hebben in alle gevallen onze specifieke vragen over de opzet van de onderzoeken zo goed mogelijk geprobeerd te beantwoorden en ons geholpen aan extra materiaal. Daarnaast bedanken we Gert-Jan Swaving van de Nationaal Coördinator Groningen voor feedback op een eerste versie van dit rapport en Angelique van Wingerden van de provincie Groningen voor feedback op een tweede versie. Tenslotte bedanken we Ans Bekkering van de Nationaal Coördinator Groningen voor het gestelde vertrouwen in ons als wetenschappelijk team, het organiseren van diverse overleggen en het begeleiden van de voortgang van dit onderzoeksproject.

**Opdrachtgever:**

Nationaal Coördinator Groningen (NCG)

**Uitvoerder/Wetenschappelijk team:**

Dr. Peter Lugtig (Universiteit Utrecht)  
 Dr. Emma ter Mors (Universiteit Leiden)  
 Prof. dr. Jelke Bethlehem (Universiteit Leiden)

**Stuurgroep:**

De stuurgroep die het onderzoek begeleidde bestond uit Mevrouw Milou Dubois (projectleider Meerjarenprogramma NCG), Mevrouw Rika Pot (voormalig burgemeester van Appingedam en lid van de Adviesraad van het Sociaal Planbureau Groningen), De heer Jan Boer en De heer Meindert Schollema (beiden Vereniging Groninger Dorpen), De heer André van de Nadort (burgemeester Ten Boer en vertegenwoordiger van de Veiligheidsregio Groningen), De heer Huub Hanssen (provincie Groningen), De heer Derwin Schorren (Vereniging Groninger Bodembeweging), en De heer Joost Kappelhof en De heer Sander Steijn (beiden Sociaal Cultureel Planbureau).

**Contactgegevens Universiteit Leiden/Universiteit Utrecht:**

Universiteit Leiden, afdeling Sociale en Organisatie Psychologie  
 Wassenaarseweg 52 T 071 527 4075  
 Postbus 9555 E-mail: emors@fsw.leidenuniv.nl  
 2300 RB Leiden

Universiteit Utrecht, afdeling Methodologie en Statistiek  
 Padualaan 14 T 030 253 4438  
 Postbus 80.140 E-mail: p.lugtig@uu.nl  
 3508 TC Utrecht

# Inhoudsopgave

1.	Samenvatting	4
2.	Inleiding	6
	2.1 Achtergrond	6
	2.2 Doelstelling onderzoek	7
	2.3 Opzet rapport	8
3.	Onderzoeksopzet	9
	3.1 Opzet en werkwijze	9
	3.2 Methoden	12
	3.3 De methodologische meetlat	13
	3.4 Relevantie voor toekomstige effectmetingen naar leefbaarheid	25
4.	Onderzoeksresultaten	28
	4.1 Gronings Perspectief (RUG)	29
	4.2 Wonen en aardbevingen in Groningen (OTB Delft)	35
	4.3 Leefbaarometer 2.0/WoON	39
	4.4 Leefbaarheid en bevolkingskrimp in Groningen (Sociaal Planbureau Groningen)	44
	4.5 Onderzoek naar de tevredenheid met de schadeafhandeling door de NAM (APE Public Economics)	48
5.	Conclusies	53
6.	Aanbevelingen	59
7.	Referentielijst	65
	Bijlage 1: Offerte aanvraag (uitnodiging) NCG/EZ	66
	Bijlage 2: Achtergrond wetenschappelijk team	70
	Bijlage 3: Longlist en shortlist survey-onderzoeken	72
	Bijlage 4: Een technisch overzicht van het trekken van steekproeven, en correctiemethodes voor non-respons door middel van weging	75
	Bijlage 5: Overzicht van onderzoeken naar leefbaarheid in aardbevingsgebied Groningen	84
	Bijlage 6: Reactie onderzoekers	87

## 1.

# Samenvatting

Dit onderzoeksrapport doet verslag van het onderzoeksproject “Het meten van de gevolgen van aardbevingen op de leefbaarheid in Groningen: Een methodologische review naar leefbaarheidsonderzoek” dat in de periode oktober 2016 – juli 2017 is uitgevoerd door de Universiteit Leiden en de Universiteit Utrecht in opdracht van de Nationaal Coördinator Groningen (NCG).

De aardgaswinning in de provincie Groningen veroorzaakt aardbevingen. Deze aardbevingen vormen een bedreiging voor de leefbaarheid in het gebied. Diverse partijen, waaronder de NCG, willen daarom voor de komende jaren op basis van de resultaten van sociaalwetenschappelijk onderzoek onder inwoners beleids- en investeringsbeslissingen nemen die bijdragen aan het verbeteren (en monitoren) van de feitelijke en waargenomen leefbaarheid in Groningen. Een belemmerende factor hierbij is de verscheidenheid aan beschikbare studies. Het is lastig om de kwaliteit en toepasbaarheid van de resultaten van de verschillende studies te beoordelen. Ook vraagt de NCG zich af hoe de effecten van gaswinning op leefbaarheid in het aardbevingsgebied in de toekomst kunnen worden gemeten.

Het huidige onderzoek richt zich op het beantwoorden van de vraag van de NCG en heeft een tweeledige doelstelling, namelijk 1) te beoordelen in hoeverre bestaande survey studies onder inwoners naar leefbaarheid kunnen dienen als basis voor het maken van beleids- en investeringsbeslissingen en 2) aanbevelingen te geven over manieren waarop toekomstig onderzoek tegemoet kan komen aan geconstateerde zwaktes en moeilijkheden van bestaand onderzoek. Deze aanbevelingen zijn erop gericht om in de toekomst goede effectmetingen te kunnen doen naar veranderingen in de leefbaarheid in het aardbevingsgebied, rekening houdend met veranderend beleid en andere maatschappelijke veranderingen (bijvoorbeeld bevolkingskrimp).

Het wetenschappelijk team heeft allereerst gekeken naar welk survey-onderzoek er tot op heden is gedaan binnen onderzoek naar leefbaarheid onder inwoners in Groningen. Vervolgens zijn er samen met de stuurgroep die dit project heeft begeleid vijf surveys gekozen om te reviewen. Het wetenschappelijk team heeft alle vijf de surveys langs een methodologische meetlat gelegd om op een rij te zetten wat de voor- en nadelen zijn van de onderzoeksopzetten waarvoor gekozen is. De methodologische meetlat die we hiervoor gebruikt hebben is opgenomen in dit rapport en kan ook in de toekomst door bestuurders en andere geïnteresseerden en belanghebbenden gebruikt worden om de kwaliteit van survey-onderzoek te beoordelen.

De conclusie van onze review is dat alle vijf de gereviewde survey studies in algemene zin goed zijn opgezet en gezamenlijk adequaat zijn om een goed beeld te geven van de leefbaarheid en veiligheidsbeleving van mensen in en buiten het aardbevingsgebied in Groningen, en dus gebruikt kunnen worden voor het maken van beleids- en investeringsbeslissingen. Dat wil niet zeggen dat er geen enkele kanttekening te plaatsen is bij de onderzoeken; zo is de hoge non-respons een probleem en is er ook op het gebied steekproefopzet en statistische correctiemethodes nog verbetering mogelijk. Daarnaast zijn er ook specifieke sterke kanten aan de verschillende methodologische opzetten van de onderzoeken. Deze sterke punten belichten we in dit rapport.

Aan de hand van deze review van vijf studies naar leefbaarheid geven we in dit rapport concreet advies aan de NCG hoe toekomstig onderzoek naar leefbaarheid in en rond het aardbevingsgebied zou kunnen worden opgezet. Op deze wijze beoogt het huidige project naast het bieden van meer inzicht in de bruikbaarheid van bestaande surveys, ook concrete inzichten en aanbevelingen op te leveren voor toekomstig survey-onderzoek naar leefbaarheid in Groningen.

## 2. Inleiding

Dit onderzoeksrapport doet verslag van het onderzoeksproject “Het meten van de gevolgen van aardbevingen op de leefbaarheid in Groningen: Een methodologische review naar leefbaarheidsonderzoeken” dat in de periode oktober 2016 – juli 2017 door onderzoekers van de Universiteit Leiden en de Universiteit Utrecht is uitgevoerd in opdracht van de Nationaal Coördinator Groningen.

### 2.1 Achtergrond

Op 16 augustus 2012 vond er nabij het Groningse dorp Huizinge een aardbeving met een geschatte sterkte van 3.6 op de schaal van Richter plaats. Dit was de krachtigste aardbeving die tot dan toe gemeten was in Groningen. De aardbeving veroorzaakte veel onrust en resulteerde in veel schademeldingen. De aardbeving in Huizinge werd gelinkt aan aardgaswinning door de Nederlandse Aardolie Maatschappij (NAM) in Groningen en kan gezien worden als het startpunt van een brede maatschappelijke discussie over gaswinning in Nederland.

Een centraal onderwerp in deze lopende discussie betreft de gevolgen van de gaswinning voor de inwoners van Groningen. Belangrijke thema's die in de periode van 2012-heden naar voren zijn gekomen zijn onder andere a) de impact van de gaswinning op de waargenomen en objectieve veiligheid van woningen en inwoners (o.a. versterking van gebouwen, in kaart brengen van risico's, beperking volume gaswinning), b) de (im)materiële schade resulterend van gaswinning (o.a. wat is schade, wie hebben er schade en om welke schade gaat het, hoe dient deze afgehandeld en voorkomen te worden, door wie, op welke wijze), c) de impact van de gaswinning op de woningmarkt (o.a. impact op waarde en courantheid van woningen, woonplezier van inwoners, verhuisbehoefte en verhuismogelijkheden) en d) in bredere zin de impact van de gaswinning en aardbevingen op leefbaarheid binnen de regio, ook gelinkt aan krimp ('Leefbaar en kansrijk Groningen', wat is een eerlijke verdeling van kosten en baten, discussies over procedurele en distributieve rechtvaardigheid, vertrouwen en wantrouwen in partijen).

In de periode van 2012-heden zijn als onderdeel en resultaat van deze maatschappelijke discussie diverse politieke besluiten genomen, nieuwe instanties en overlegorganen opgericht, en (al dan niet tijdelijke) maatregelen ingesteld om de negatieve impact van gaswinning voor de inwoners van Groningen te verminderen of voorkomen, en om een toename van positieve uitkomsten voor de regio te bewerkstelligen. Diverse partijen willen voor de komende jaren op basis van de resultaten van studies beleids- en investeringsbeslissingen nemen die bijdragen aan het verbeteren (en monitoren) van de feitelijke en waargenomen veiligheid en leefbaarheid van Groningen. Binnen het aardbevingendossier zijn er de afgelopen jaren meerdere kwantitatieve (bijvoorbeeld vragenlijstonderzoek) en kwalitatieve (bijvoorbeeld: interviews, focusgroepen, versnellingsstafels) onderzoeken onder inwoners van Groningen uitgevoerd over uiteenlopende onderwerpen (voor een indicatief overzicht, zie [Bijlage 5](#)). Het gaat hierbij onder andere om studies die (on)tevredenheid over de afhandeling van schades in kaart brengen en studies die tevredenheid met de woonomgeving en factoren die hierbij bepalend zijn onderzoeken. Een belemmerende factor voor partijen die beleids- en investeringsbeslissingen willen nemen is echter de verscheidenheid in beschikbare

studies—zowel verscheidenheid qua opzet, focus en resultaten. Deze verscheidenheid maakt het lastig om de kwaliteit en toepasbaarheid van de resultaten van de verschillende studies te beoordelen. Een concrete vraag die daarnaast bij de Nationaal Coördinator Groningen (NCG) leeft is hoe nieuw of vervolgonderzoek naar de effecten van gaswinning op de leefbaarheid in het aardbevingsgebied er concreet uit zou moeten zien.

## 2.2 Doelstelling onderzoek

De huidige methodologische review richt zich op de vraag van de NCG hoe de effecten van gaswinning en de daardoor veroorzaakte aardbevingen op leefbaarheid in het aardbevingsgebied in de toekomst kunnen worden gemeten. We richten ons hierbij, op verzoek van de NCG, op kwantitatief survey-onderzoek onder inwoners van het aardbevingsgebied, en daar omheen liggende gebieden. Er is de afgelopen jaren ook kwalitatief onderzoek (zoals interviews, focusgroepen, versnellingsstafels) uitgevoerd, soms in combinatie met survey-onderzoek, maar die kwalitatieve onderzoeken reviewen we hier niet.

Om de vraag van de NCG met betrekking tot de opzet van effectmetingen te beantwoorden kijken wij naar welk survey-onderzoek er tot op heden is gedaan binnen onderzoek naar leefbaarheid. Dit doen we door vijf bestaande studies binnen het aardbevingendossier te reviewen: Dat wil zeggen dat we de onderzoeken langs een methodologische meetlat leggen. Het doel daarvan is op een rij te zetten wat de voor- en nadelen zijn van de onderzoeksopzetten waarvoor gekozen is, en aan de hand daarvan een advies te geven aan de NCG over de opzet van toekomstig onderzoek naar leefbaarheid. Op deze wijze beoogt het huidige project naast het bieden van meer inzicht in de bruikbaarheid van bestaande surveys voor beleids- en investeringsbeslissingen, ook concrete inzichten en aanbevelingen op te leveren voor toekomstig survey-onderzoek naar leefbaarheid in Groningen. Welke vijf bestaande studies zijn gereviewd wordt uiteengezet in [Hoofdstuk 3 \(“Onderzoeksopzet”\)](#). Hier wordt ook de methodologische meetlat geïntroduceerd en uitgelegd.

De doelstelling van het onderhavige onderzoek “Het meten van de gevolgen van aardbevingen op de leefbaarheid in Groningen: een methodologische review naar leefbaarheidsonderzoeken” is kortom om 1) te beoordelen in hoeverre bestaande studies kunnen dienen als basis voor het maken van beleids- en investeringsbeslissingen en 2) aanbevelingen te geven over de manieren waarop toekomstig onderzoek tegemoet kan komen aan geconstateerde zwaktes en moeilijkheden van bestaand onderzoek. Deze aanbevelingen zijn erop gericht om in de toekomst goede effectmetingen te kunnen doen naar veranderingen in de leefbaarheid in het aardbevingsgebied, rekening houdend met veranderend beleid en andere maatschappelijke veranderingen (bijvoorbeeld bevolkingskrimp).

Het begrip leefbaarheid wordt in de bestuurlijke context in Nederland veel gebruikt. Wat partijen onder leefbaarheid verstaan, is echter niet altijd helder of eenduidig. Leefbaarheid in enge zin gedefinieerd gaat puur om de tevredenheid met de woonomgeving; bij een bredere definitie van leefbaarheid worden daarentegen ook andere aspecten meegenomen ([Knol, Maas, & Roes, 2002](#)). Onder het begrip leefbaarheid vallen dan bijvoorbeeld de ervaren fysieke kwaliteit van de woonomgeving (het huis zelf, en de

omgeving daarvan), de sociale kwaliteit (bijvoorbeeld: tevredenheid met voorzieningen, omgang met mensen uit de buurt), sociale kenmerken (bijvoorbeeld: beschikbaarheid van voorzieningen, bevolkingsopbouw), en de veiligheid van de woonomgeving. Wij zullen in dit rapport proberen aan te sluiten bij deze brede definitie van leefbaarheid.

## 2.3 Opzet rapport

Dit conceptrapport kent de volgende opzet. In Hoofdstuk 3 “Onderzoeksopzet” wordt de onderzoeksmethode beschreven en verantwoord. In Hoofdstuk 4 “Onderzoeksresultaten” worden de onderzoeksresultaten weergegeven. In Hoofdstuk 5 “Conclusies” worden de bevindingen uit Hoofdstuk 4 geïntegreerd besproken. Tenslotte wordt in Hoofdstuk 6 “Aanbevelingen” een aantal aanbevelingen gedaan ten behoeve van toekomstig survey-onderzoek naar leefbaarheid in Groningen.

## 3. Onderzoeksopzet

Dit hoofdstuk doet verslag van de onderzoeksopzet van de methodologische review die in de periode oktober 2016 – juli 2017 door de Universiteit Leiden en de Universiteit Utrecht is uitgevoerd in opdracht van de Nationaal Coördinator Groningen (NCG).

### 3.1 Opzet en werkwijze

#### Wetenschappelijk team

Het onafhankelijke wetenschappelijk team dat het onderzoek opzette, uitvoerde en rapporteerde bestond uit Emma ter Mors en Jelke Bethlehem van de Universiteit Leiden en Peter Lugtig van de Universiteit Utrecht. Het wetenschappelijk team werd ondersteund door Jildou Talman (studentassistente Universiteit Leiden). In [Bijlage 2](#) is een beknopt Curriculum Vitae van de leden van het wetenschappelijk team opgenomen. De onderzoekers van de Universiteit Leiden en Universiteit Utrecht hebben allereerst de onderzoeksopzet uitgewerkt. Vervolgens is overleg gevoerd met de NCG als opdrachtgever en met de stuurgroep die de review begeleidde omtrent vraagstelling en onderzoeksaanpak. Gedurende het onderzoek hebben de Universiteit Leiden en de Universiteit Utrecht het onderzoek gecoördineerd en de onderzoeksgegevens verzameld, geanalyseerd en gerapporteerd.

#### Stuurgroep

Om de onafhankelijkheid van de review te garanderen en zo ook een breed maatschappelijk draagvlak voor de resultaten van het onderzoek te borgen, vonden zowel de NCG als het wetenschappelijk team het belangrijk dat er naast een onafhankelijk wetenschappelijk reviewteam ook een stuurgroep met diverse achtergrond zou worden ingesteld binnen het reviewproject. De stuurgroep die het onderzoek begeleidde bestond uit Mevrouw Milou Dubois (projectleider Meerjarenprogramma NCG), Mevrouw Rika Pot (voormalig burgemeester van Appingedam en lid van de Adviesraad van het Sociaal Planbureau Groningen), De heer Jan Boer en De heer Meindert Schollema (beiden Vereniging Groninger Dorpen), De heer André van de Nadort (burgemeester Ten Boer en vertegenwoordiger van de Veiligheidsregio Groningen), De heer Huub Hanssen (provincie Groningen), De heer Derwin Schorren (Vereniging Groninger Bodembeweging), De heer Joost Kappelhof en De heer Sander Steijn (beiden Sociaal Cultureel Planbureau). De taak van de stuurgroep bestond uit het becommentariëren van de onderzoeksaanpak, de selectie van de te reviewen onderzoeken, en het lezen en becommentariëren van een voorlopige versie van dit rapport<sup>1</sup>. Wij danken hen allen voor hun feedback en bijdrage in de totstandkoming van dit rapport.

#### Onderzoeksaanpak

De basis voor dit onderzoek vormde de door NCG uitgeschreven onderzoeksopdracht welke is opgenomen als [Bijlage 1](#) en de door het wetenschappelijke team geschreven offerte.

Na toekenning van de opdracht is het wetenschappelijk team in oktober 2016 in overleg met de NCG begonnen met het samenstellen van een stuurgroep. Op basis van een door het NCG opgestelde longlist van onderzoeken die relevant zijn, heeft

---

<sup>1</sup> Alle leden van de stuurgroep hebben feedback gegeven op (delen van) dit rapport. Derwin Schorren heeft geen stukken kunnen becommentariëren.

het wetenschappelijk team in januari 2017 een shortlist van vier survey-onderzoeken opgesteld om te reviewen (zie [Bijlage 3](#) voor de longlist en de shortlist). De criteria die we daarbij hebben gebruikt zijn:

1. Elk onderzoek is uitgevoerd door een andere partij.
2. Het onderzoek is van een voldoende omvang om inzicht te geven in de leefbaarheid in het aardbevingsgebied in Groningen.
3. Er is variatie in de onderwerpen die in de onderzoeken aan bod komen. Als gemene deler hebben alle onderzoeken het thema 'leefbaarheid', maar er is bewust voor gekozen voor variatie in de andere thema's die in de onderzoeken aan bod komen (denk aan tevredenheid met schadeafhandeling; de impact van aardbevingen op de woningmarkt; de impact van aardbevingen op de gezondheid van inwoners).
4. Er is gekozen voor onderzoeken die verschillen in de gehanteerde methodologie. Het onderzoeksteam heeft bewust gezocht naar tenminste één onderzoek dat longitudinaal is uitgevoerd, en één onderzoek dat landelijk is uitgevoerd.

Het is zowel bij het samenstellen van de longlist als de shortlist nadrukkelijk niet de bedoeling geweest om uitputtend te werk te gaan. In [Hoofdstuk 2](#) hebben we al aangegeven dat er de afgelopen jaren diverse onderzoeken naar de impact van aardbevingen zijn uitgevoerd onder inwoners van Groningen. De geselecteerde survey-onderzoeken zijn dientengevolge niet representatief voor al het onderzoek dat is uitgevoerd. Het survey-onderzoek dat de longlist niet heeft gehaald is veelal kleinschaliger en meer ad-hoc van opzet. In algemene zin geldt dat de methodologische kwaliteit van dat soort onderzoeken vaak lager is dan van landelijke, grootschalige en longitudinale studies. Dat wil echter niet zeggen dat niet-opgenomen studies van lage kwaliteit zijn. Tijdens het onderzoeksproces zijn we bijvoorbeeld meerdere keren gestuit op resultaten van de 'monitor Eemsdelta', een onderzoek dat naar onze mening interessant is voor het begrijpen van de leefbaarheid in dit gebied. Ook bestaat er relatief veel kwalitatief onderzoek binnen het dossier, bijvoorbeeld in de vorm van interviews, dat een essentiële bijdrage levert aan het begrijpen van de specifieke problemen die inwoners in Groningen ervaren als het gaat om leefbaarheid.

Tijdens een eerste vergadering met de stuurgroep op 19 januari 2017 is de longlist, alsmede de shortlist met de voorkeur van NCG en het wetenschappelijk team, voorgelegd aan de stuurgroep. Op basis van een discussie met de stuurgroep is er uiteindelijk gekozen voor het reviewen van de volgende 5 studies:

- 'Gronings Perspectief', uitgevoerd door de Rijksuniversiteit Groningen in samenwerking met de GGD en Onderzoeksbureau Onderzoek en Statistiek Groningen.
- 'Wonen en aardbevingen in Groningen', uitgevoerd door OTB Delft in samenwerking met CMO STAMM. Dit is een deelrapport van het grotere onderzoek 'Woningmarkt en leefbaarheidsonderzoek aardbevingsgebied Groningen'.
- 'De Leefbaarometer 2.0', uitgevoerd door de Atlas van Gemeenten en RIGO Research en Advies.
- Het onderzoek 'Leefbaarheid en bevolkingskrimp in Groningen', uitgevoerd met behulp van het Groninger Panel door het Sociaal Planbureau Groningen, onderdeel van onderzoeksbureau CMO STAMM.
- 'Onderzoek naar de tevredenheid met de schadeafhandeling door de NAM' door onderzoeksbureau APE Public Economics.

We verwijzen naar [Bijlage 5](#) voor details over het door ons geraadpleegde materiaal. We hebben ervoor gekozen in dit rapport zo min mogelijk te refereren aan specifieke passages uit de geraadpleegde materiaal, maar de bovenstaande onderzoeken als geheel te bespreken.

Behalve over de selectie van de onderzoeken is er met de stuurgroep gesproken over de aanpak en de focus van de review (in hoeverre er aandacht moest zijn voor onder andere de kwaliteit van vraagstelling, kwaliteit van de steekproef, dekkingsfouten, non-respons, kwaliteit van analyses en conclusies), de mate van detail van dit rapport, en het type aanbevelingen dat wenselijk zou zijn als uitkomst van ons onderzoek. Besloten is om alle vijf de onderzoeken langs dezelfde methodologische meetlat te leggen. Deze methodologische meetlat wordt in [sectie 3.3](#) uiteengezet. Daarnaast is besloten om in het onderzoeksrapport voor alle vijf de onderzoeken de relevantie voor toekomstige effectmetingen naar leefbaarheid te bespreken (zie [sectie 3.4](#)).

In de periode van januari-mei 2017 heeft het wetenschappelijk team informatie verzameld over de vijf te reviewen onderzoeken. Allereerst hebben we daarbij gebruik gemaakt van publiek beschikbare rapporten en bronmateriaal van de survey-onderzoeken (zoals vragenlijsten en details over de steekproeftrekking). Op basis van het beschikbare materiaal is in een eerste ronde een voorlopige analyse gedaan van de methodologisch sterke en minder sterke kanten van ieder onderzoek. Bij alle vijf de onderzoeken gaf onze eerste analyse aanleiding voor verdiepende vragen over de exacte procedures die in de onderzoeken zijn gevolgd. De vragen verschilden per onderzoek, maar betroffen vooral details over het veldwerk (bijvoorbeeld: duur van veldwerk, non-respons, details over de getrokken steekproef en correcties voor non-respons). Naar aanleiding van deze aanvullende vragen heeft het wetenschappelijk team contact opgenomen met de onderzoekers en heeft er soms mondeling en soms schriftelijk overleg plaatsgevonden met de onderzoekers. In alle gevallen heeft het wetenschappelijk team op alle vragen een afdoende antwoord gekregen. We bedanken de onderzoekers hierbij nogmaals voor de aanvullende informatie en de prettige samenwerking.

Op basis van het publiek beschikbare materiaal en aanvullend materiaal heeft het wetenschappelijke team alle vijf de onderzoeken langs de methodologische meetlat gelegd is er gekeken naar de relevantie van de gereviewde onderzoeken voor toekomstige effectmetingen naar leefbaarheid. Op 11 mei 2017 hebben we een conceptversie van dit rapport opgeleverd aan de stuurgroep. Dit rapport hebben we vervolgens op 22 mei 2017 besproken is met de stuurgroep. Parallel aan dit traject hebben we het conceptrapport voorgelegd aan de onderzoekers van de gereviewde onderzoeken, met het verzoek om feedback op het rapport in het algemeen en op de review van het eigen onderzoek te geven.

Naar aanleiding van de feedback van de stuurgroep en de feedback van onderzoekers hebben we in de maand juni delen van dit rapport herschreven. Op 21 juni 2017 is een tweede versie van dit rapport gedeeld met de stuurgroep en de onderzoekers van de gereviewde studies. Op 28 juni 2017 is de tweede versie van het conceptrapport besproken met de stuurgroep. We hebben de feedback van de stuurgroep en de feedback van de onderzoekers vervolgens verwerkt, wat heeft geresulteerd in het opleveren van het huidige eindrapport op 7 juli 2017. We hebben in het kader van hoorwederhoor de onderzoekers van de vijf gereviewde onderzoeken ook de mogelijkheid geboden een schriftelijke inhoudelijke reactie te geven op ons rapport. Ze konden hierin

bijvoorbeeld extra informatie of uitleg geven over de eigen onderzoeks aanpak, reageren op de bevindingen en conclusies van onze review, of anderszins een reactie geven. De afspraak hierbij was dat we de reactie integraal zouden opnemen in het eindrapport. Eén van de vijf betrokken onderzoeksteams heeft voor de mogelijkheid gekozen. We hebben de reactie van dit onderzoeksteam opgenomen in [Bijlage 6](#).

In dit rapport gaan we in op de kenmerken van elk gereviewde studie en de sterktes en zwaktes van de gebruikte methodologische opzet. Het doel van onze analyse is tweeledig: 1) beoordelen in hoeverre bestaande studies kunnen dienen als basis voor het maken van beleids- en investeringsbeslissingen en 2) het geven van aanbevelingen over de manieren waarop toekomstig onderzoek tegemoet kan komen aan geconstateerde zwaktes en moeilijkheden van bestaand onderzoek. Deze aanbevelingen zijn erop gericht om in de toekomst goede effectmetingen te kunnen doen naar veranderingen in de leefbaarheid in het aardbevingsgebied, rekening houdend met veranderend beleid en andere maatschappelijke veranderingen (bijvoorbeeld bevolkingskrimp).

In dit rapport laten we ook belangrijke aspecten buiten beschouwing. Zo doen we geen uitspraken over de leefbaarheid in het aardbevingsgebied, of veranderingen daarin sinds 2012. Ook zullen we de uitkomsten van de gereviewde onderzoeken niet expliciet proberen samen te vatten of vergelijken. Een van de conclusies van dit rapport zal zijn dat de gereviewde onderzoeken qua inhoud en methodologische opzet erg van elkaar verschillen. Daardoor zijn de uitkomsten van de onderzoeken niet makkelijk te vergelijken, maar moeten de uitkomsten van de onderzoeken vooral als aanvullend op elkaar worden gezien.

## 3.2 Methoden

Peilingen, enquêtes, surveys. Het zijn verschillende namen voor hetzelfde type onderzoek. In dit rapport reviewen we vragenlijstonderzoeken naar de leefbaarheid in Groningen. Deze onderzoeken worden in dit rapport aangeduid met de term *surveys*.

Survey-onderzoeken zijn vooral geschikt voor het in kaart brengen van de meningen en gevoelens van een relatief grote groep mensen. Voor het uitvoeren van een survey bestaan gestandaardiseerde procedures, die achtereenvolgend worden doorlopen: Allereerst wordt uit een populatie van personen of huishoudens een steekproef getrokken. Aan de geselecteerde personen wordt gevraagd een vragenlijst in te vullen. De zo verzamelde gegevens worden geanalyseerd en opgeschreven. Als het onderzoek goed is opgezet en uitgevoerd, dan kunnen de conclusies op verantwoorde wijze worden vertaald van de steekproef naar de gehele populatie. Voldoet het onderzoek niet aan de eisen, dan is die generalisatie van steekproef naar populatie niet mogelijk.

Ook worden surveys gebruikt om verschillen in meningen te verbinden aan kenmerken van mensen. Bijvoorbeeld: zijn onveiligheidsgevoelens hoger bij mannen of vrouwen? Binnen surveys worden vooral gesloten vragen gesteld, met antwoordcategorieën die vooraf door de onderzoeker bepaald zijn. Daardoor zijn surveys minder geschikt om precies te begrijpen *waarom* een persoon een bepaalde mening heeft of zich bijvoorbeeld onveilig voelt. Daarvoor kunnen beter kwalitatieve interviews worden gebruikt.

De manier waarop een survey wordt opgezet en uitgevoerd is in hoge mate gestandaardiseerd, wat het mogelijk maakt om iets te zeggen over de relatieve sterktes en zwaktes van surveys, onafhankelijk van het onderwerp. Om te kunnen vaststellen of surveys van voldoende kwaliteit zijn, wordt in dit onderzoek gebruik gemaakt van een methodologische meetlat. De oorspronkelijk versie van deze meetlat 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). Zie ook [Bethlehem \(2012\)](#). De hier gehanteerde versie van de meetlat is een aangepaste versie van deze oorspronkelijke meetlat. De oorspronkelijke meetlat was bedoeld voor journalisten, opdat zij kunnen beoordelen of een vragenlijstonderzoek serieus is opgezet of niet. In onze meetlat willen we nadrukkelijk proberen nuance aan te brengen. Vaak zijn methodologische keuzes namelijk niet simpelweg goed of fout, maar afhankelijk van de context van het onderzoek passend of minder passend.

Aanleiding voor het opstellen van de oorspronkelijke meetlat was de grote toename van het aantal surveys en peilingen. Het internet maakt het mogelijk eenvoudig, snel en goedkoop bij heel veel mensen informatie te verzamelen. De vraag is echter of al die surveys en peilingen wel een goed beeld geven van de werkelijkheid. Door de criteria in de meetlat een voor een af te lopen, kan een indruk worden verkregen van de kwaliteit van een onderzoek, en de toepasbaarheid daarvan voor bijvoorbeeld het maken van beleids- en investeringsbeslissingen. Dit doen we in de huidige review.

In [sectie 3.3](#) wordt de meetlat beschreven die zal helpen ons een antwoord te geven op onze eerste onderzoeksvraag: In hoeverre bestaande studies kunnen dienen als basis voor het maken van beleids- en investeringsbeslissingen. We kijken in onze review onder andere naar de kwaliteit van de onderzoeksofzet, uitvoering, rapportage en analyses. De meetlat wordt in de volgende sectie puntsgewijs simpel uiteengezet (zie blz. 13-15). Daarna volgt per punt van de meetlat ter verduidelijking en verdieping verdere uitleg en duiding (zie blz. 15-25). Technische verhandelingen en jargon hebben we hierbij zoveel mogelijk vermeden om het rapport en de review helder en toegankelijk te houden. Voor de geïnteresseerde lezer hebben we in [Bijlage 4](#) van dit rapport aanvullende, meer technische informatie behorend bij de meetlat opgenomen (o.a. extra informatie over het trekken van steekproeven en het corrigeren voor non-respons).

In [sectie 3.4](#) (blz. 25 en verder) bespreken we tenslotte op welke manier we tot een antwoord proberen te komen op onze tweede onderzoeksvraag: “Op welke manier kan toekomstig onderzoek tegemoetkomen aan geconstateerde zwaktes en moeilijkheden van bestaand onderzoek? Aanbevelingen die hieruit voortvloeien moeten erop gericht zijn om in de toekomst goede effectmetingen te kunnen doen naar veranderingen in de leefbaarheid in het aardbevingsgebied, rekening houdend met veranderend beleid en andere maatschappelijke veranderingen (bijvoorbeeld bevolkingskrimp)”.

### 3.3 De methodologische meetlat

Onze methodologische meetlat bestaat uit negen onderdelen: acht van die onderdelen hebben betrekking op de manier waarop data is verzameld. Het negende onderdeel betreft de manier waarop deze data worden gebruikt om een antwoord te geven op een onderzoeksvraag.

- **De onderzoeksverantwoording:**  
Goede survey-onderzoeken gaan gepaard met een document waarin precies staat uitgelegd hoe de survey is opgezet en uitgevoerd. De mate van detail en de helderheid van de onderzoeksverantwoording bepaalt hoe repliceerbaar en transparant een onderzoek is, en bepaalt daarmee ook deels de kwaliteit van het onderzoek zelf.
- **De opdrachtgever en/of financier:**  
Wanneer de opdrachtgever en/of financier geen duidelijk belang heeft bij een specifieke onderzoeksuitkomst, dan waarborgt dat de onafhankelijkheid en neutraliteit van het uitgevoerde onderzoek, en daarmee deels de kwaliteit van het onderzoek.
- **De doelpopulatie:**  
In goede survey-onderzoeken staat helder beschreven naar welke populatie de uitkomsten dienen te worden gegeneraliseerd.
- **De steekproef:**  
Een goede steekproef is willekeurig (aselect) getrokken uit de doelpopulatie.
- **Het responspercentage en selectiviteit van de respons:**  
De grootte van de getrokken bruto steekproef en de hoeveelheid respondenten die uiteindelijk de vragenlijst invult, bepalen het responspercentage. Een kwalitatief goed onderzoek wordt gekenmerkt door een hoge respons. Ook wordt in goed survey-onderzoek de non-respons geanalyseerd op selectiviteit.
- **De correctie voor non-respons:**  
In de praktijk zijn er altijd mensen die wel geselecteerd zijn in de steekproef, maar niet meedoen aan het onderzoek. Wanneer er selectiviteit is in die respons kan daarvoor worden gecorrigeerd door middel van weegtechnieken. Goede survey-onderzoeken corrigeren voor non-respons, en gebruiken daarvoor een geschikt weegmodel.
- **Onzekerheidsmarges:**  
Doordat een steekproef wordt getrokken uit een veel grotere populatie bestaat er altijd statistische onzekerheid over de uitkomsten. Goede onderzoeken vermelden die onzekerheidsmarges, en bespreken de implicaties daarvan voor de uitkomsten van het onderzoek.
- **De vragenlijst:**  
Goede vragenlijsten zijn volledig (dekken alle onderwerpen die van belang zijn voor het beantwoorden van de onderzoeksvraag). De vragen in de vragenlijst en individuele vragen zijn helder, niet sturend, en voor iedereen eenduidig te begrijpen.
- **De kwaliteit van de analyses:**  
Bovenstaande acht onderdelen van de meetlat hebben betrekking op de kwaliteit van de opzet en uitvoering van het onderzoek. Ze beschrijven het proces van het verzamelen van antwoorden op vragen onder een kleine groep mensen uit een grotere populatie. Nadat deze data zijn verzameld worden er analyses gedaan om een antwoord te krijgen op de gestelde onderzoeksvragen. Er zijn hierbij verschillende analyses mogelijk. Analyses kunnen bijvoorbeeld beschrijvend van aard zijn (zoals frequentietabellen over de onveiligheidsgevoelens van de respondenten), maar ook gericht zijn op het onderzoeken en toetsen van relaties tussen variabelen (bijvoorbeeld: Zijn mensen in het aardbevingsgebied minder tevreden over hun woonomgeving dan mensen buiten het aardbevingsgebied). In deze analyses kunnen fouten worden gemaakt. In goed onderzoek worden de juiste analysemethodes

gebruikt om antwoord te geven op de onderzoeksvragen.

In de rest van deze sectie wordt de meetlat op toegankelijke wijze nader toegelicht en wordt verder uitgelegd aan welke criteria goed survey-onderzoek voldoet. Voor aanvullende, meer technische informatie behorende bij de meetlat verwijzen we naar [Bijlage 4](#) van dit rapport.

### 3.3.1 De onderzoeksverantwoording

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

De onderzoeksverantwoording moet minimaal de volgende aspecten bevatten en beschrijven:

- De namen van de opdrachtgever en de financier.
- De namen van de organisaties die de survey hebben opgezet en uitgevoerd.
- De doelpopulatie:  
Dat is de groep waar de uitkomsten van het onderzoek betrekking op hebben.
- De volledige vragenlijst.  
Ook dient duidelijk te worden gemaakt of de vragenlijst van tevoren is getest bij respondenten om de kwaliteit van de vragen en de responsen te verhogen.
- Het gebruikte steekproefkader:  
Dit is de lijst waarin voor ieder lid van de doelpopulatie de contactgegevens staan (adres, telefoonnummer, e-mailadres, e.d.) en waaruit de steekproef is getrokken. De lijst zelf kan nooit in een rapport worden opgenomen in verband met het waarborgen van de anonimiteit van respondenten. Wat wel aanwezig moet zijn is een heldere beschrijving van de inhoud van de lijst.
- De wijze waarop de steekproef is getrokken.  
Er moet duidelijk worden aangegeven op welke manier de steekproef uit het steekproefkader (de lijst) is geselecteerd. Bij een goede steekproef is voor elke persoon uit de populatie te herleiden wat (bij benadering) de kans was om in de steekproef terecht te komen. Soms wordt ervoor gekozen om bepaalde bevolkingsgroepen een hogere kans te geven om in de steekproef terecht te komen. Dit kan bijvoorbeeld gebeuren om voldoende respondenten te krijgen uit een kleine, maar voor het onderzoek interessante bevolkingsgroep (bijvoorbeeld bewoners van monumentale panden in het kader van het huidige reviewproject naar de impact van aardbevingen op de leefbaarheid in Groningen). Als de steekproef uit een online panel is getrokken, moet duidelijk zijn hoe dat panel tot stand is gekomen.
- De omvang van de bruto steekproef.  
Dit is de omvang van de initiële steekproef zoals deze is getrokken uit het steekproefkader.
- De omvang van de respons (netto steekproef).  
Dit is het aantal respondenten, dat wil zeggen het aantal personen dat uitgenodigd is en de vragen uit de survey beantwoord heeft.
- Het percentage respons ( $100 \times \text{Netto steekproef} / \text{Bruto steekproef}$ ).
- Een beschrijving van de analyse van non-respons.  
Op basis van gegevens die bekend zijn voor de hele populatie kan worden bekeken of er verschillen zijn tussen personen die wel hebben meegedaan

- (respondenten) en personen die niet hebben meegedaan (non-respondenten). De wijze waarop de respons is gecorrigeerd voor de negatieve effecten van non-respons (of het feit dat ongelijke kansen zijn gebruikt in de steekproeftrekking). Om hiervoor te corrigeren wordt meestal een statistische correctieprocedure gebruikt die 'wegen' wordt genoemd. Minimaal moeten de gebruikte weegvariabelen worden genoemd. Bij voorkeur wordt ook uitgelegd hoe die variabelen zijn gebruikt in de berekening van de gewichten (het weegmodel).

### 3.3.2 De opdrachtgever en/of financier

Het is van belang te weten wie de opdrachtgever en/of financier is van het onderzoek. Het komt voor dat een opdrachtgever er belang bij heeft dat de uitkomsten van een survey 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.

### 3.3.3 De doelpopulatie

De doelpopulatie is de groep van mensen waarop de uitkomsten van het onderzoek betrekking heeft. De doelpopulatie moet duidelijk worden omschreven. Doelpopulaties kunnen op allerlei manieren worden gedefinieerd. Daarbij kan sprake van geografische en demografische beperkingen. In de surveys onder de inwoners van Groningen bestaat de doelpopulatie soms uit personen uit alle gemeenten van de provincie Groningen, soms uit personen uit alleen de aardbevingsgemeenten. Er kan ook een leeftijdsgrens zijn, bijvoorbeeld alleen personen vanaf 16 of vanaf 18 jaar. En in sommige gevallen gaan onderzoeken eigenlijk niet over personen, maar over woningen. Als er bijvoorbeeld wordt gekeken naar schades aan huizen, dan bestaat de doelpopulatie niet uit personen, maar uit complete huishoudens.

De keuze van de doelpopulatie bepaalt de groep waaruit de steekproef moet worden getrokken en de populatie waarop de uitkomsten betrekking hebben. Als de doelpopulaties van verschillende surveys verschillen, kan het lastig worden om de uitkomsten van die onderzoeken met elkaar te vergelijken. Het kan bijvoorbeeld zo zijn dat de leefbaarheid alleen in bepaalde geografische delen van Groningen is veranderd door aardbevingen, of juist dat specifieke groepen mensen een verandering in leefbaarheid ervaren.

### 3.3.4 De steekproef

Een steekproef kan op allerlei manieren worden getrokken uit een doelpopulatie, maar er is maar één manier om het goed te doen en dat is met behulp van een *aselecte steekproef*. Om de uitkomsten van een survey te kunnen generaliseren naar de gehele doelpopulatie, moet de steekproef een goede afspiegeling zijn van die doelpopulatie. Daarvoor wordt ook wel de term *representativiteit* gehanteerd.

Een steekproef kan representatief worden genoemd als voor alle variabelen in het onderzoek de verdeling in de respons van de survey overeenkomt met de verdeling in de gehele doelpopulatie. Hierbij dient te worden opgemerkt dat representativiteit met betrekking tot een variabele niet automatisch betekent dat de steekproef ook

representatief is met betrekking tot andere variabelen. Het is bijvoorbeeld heel goed mogelijk de steekproef representatief te maken met betrekking tot geslacht (evenveel mannen als vrouwen), maar dan is de steekproef niet ook automatisch representatief met betrekking tot leeftijd, of–belangrijker nog–leefbaarheid.

De aangewezen manier om een steekproef te trekken die op zijn minst bij benadering representatief is met betrekking tot alle variabelen, is de *enkelvoudige aselechte steekproef* (*simple random sample*). Dat is een aselechte steekproef waarin elke persoon in de doelpopulatie dezelfde selectiekans heeft en er uit de doelpopulatie willekeurig mensen worden geselecteerd in de steekproef.

Als de steekproef op een andere wijze is geselecteerd, dan is er geen enkele garantie dat de resultaten van surveys valide (zuiver) zijn. Bij dit soort niet-aselechte steekproeven is het daarom niet mogelijk om te generaliseren van de steekproef naar de populatie.

Vooraf bij online surveys komt het vaak voor dat de steekproef niet wordt geloot, maar wordt verkregen via zelfselectie. Daarbij wordt de vragenlijst op het internet gezet en voor iedereen toegankelijk gemaakt. Via allerlei media worden mensen attent gemaakt op de survey. De hoop is dan dat men zich hierdoor laat overhalen tot invullen van de vragenlijst. Dit heet *zelfselectie*. Het zijn immers de mensen zelf die spontaan bepalen of ze meedoen aan een survey. De onderzoeker heeft daar geen controle over.

Zelfselectie heeft verschillende nadelen. Een eerste nadeel is dat iedereen kan meedoen aan het onderzoek, dus ook personen die niet tot de doelpopulatie behoren. Een tweede nadeel is dat het soms mogelijk is dat een persoon de vragenlijst meer dan één keer invult. Een derde nadeel is dat groepen mensen kunnen proberen de uitkomsten van een peiling te manipuleren. Dit alles leidt ertoe dat steekproeven op basis van zelfselectie niet representatief zijn.

Een goed voorbeeld van de effecten van zelfselectie was te zien tijdens de campagne voor de parlementsverkiezingen van 7 mei 2015 in Groot-Brittannië. De peilingen zaten er allemaal naast. Sommige Engelse media spraken van een ‘The UK polling disaster’. De peilers voorspelden allemaal een nek-aan-nek race tussen Labour en de Conservatieven. Dat zou weer een ‘hung parliament’ betekenen en dus coalitieonderhandelingen. Maar de uitslag was heel anders: De conservatieven wonnen met een comfortabele meerderheid van 6,5% en konden alleen gaan regeren. Om te kunnen vaststellen wat er mis was gegaan besloot de British Polling Council (BPC), een organisatie van opiniepeilers, een onderzoek in te stellen. De belangrijkste conclusie was dat de steekproeven van de peilers niet representatief waren. Ze waren geen goede afspiegeling van de Britse bevolking; vooral een specifieke groep oudere stemmers op de Conservatieven ontbrak. De door de peilers gebruikte correctietechnieken om hun peilingen representatief te maken, hielpen niet.

De American Association for Public Opinion Research (AAPOR) waarschuwt voor de gevaren van zelfselectie ([Baker et al., 2010](#)): *“Only when a web-based survey adheres to established principles of scientific data collection can it be characterized as representing the population from which the sample was drawn. But if it uses volunteer respondents, allows respondents to participate in the survey more than once, or excludes portions of the population from participating, it must be characterized as unscientific and is unrepresentative of any population.”* [vertaling: Dataverzameling via internet-

vragenlijsten is alleen wetenschappelijk te noemen als de steekproef de doelpopulatie vertegenwoordigt. Maar als er niet-aselecte steekproeven worden gebruikt, waarbij bepaalde groepen mensen geen kans krijgen om mee te doen, of anderen juist meer dan één keer kunnen doen, dan is de dataverzameling niet wetenschappelijk en niet representatief voor welke populatie dan ook].

Het is dus belangrijk om een aselecte steekproef te trekken. Daarvoor is een steekproefkader nodig. Dit is de (elektronische) lijst waarin voor ieder persoon in de doelpopulatie de contactgegevens staan (adres, telefoonnummer, e-mailadres, e.d.). Het steekproefkader moet precies de doelpopulatie afdekken. Als dit niet het geval is, dan kunnen verkeerde conclusies worden getrokken uit de survey. In de eerste plaats kunnen er personen in de doelpopulatie zitten die niet staan vermeld in het steekproefkader. Deze situatie heet *onderdekking*. Dit verschijnsel doet zich bijvoorbeeld voor als onderzoek wordt gedaan naar illegaal in het land verblijvende mensen terwijl de steekproef is getrokken uit het bevolkingsregister. Het kan ook gebeuren dat er personen in het steekproefkader zitten die niet thuishoren in de doelpopulatie. Dat verschijnsel wordt aangeduid met *overdekking*. Het ten onrechte meenemen van deze personen in het onderzoek kan ook tot verkeerde conclusies leiden. Overdekking kan vrij eenvoudig worden ontdekt in het veld, door voor elke persoon in de steekproef eerst te controleren of hij/zij wel voldoet aan de definitie van de doelpopulatie.

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.

Voor elke survey moet duidelijk worden gedocumenteerd uit welk steekproefkader de steekproef is getrokken. Ook moet duidelijk worden gemaakt of zich bij het trekken van de steekproef uit het steekproefkader problemen hebben voorgedaan.

Bij het uitvoeren van een survey is er sprake van een bruto steekproef en een netto steekproef. De *bruto steekproef* (de initiële steekproef) is de omvang van de steekproef die uit het steekproefkader is getrokken. In de praktijk doen zich meestal allerlei problemen voor waardoor niet alle geselecteerde respondenten de vragenlijst invullen. Dit wordt *non-respons* genoemd. Daardoor zal de uiteindelijke *netto steekproef* (de respons) kleiner zijn dan de bruto steekproef. De omvang van de netto steekproef bepaalt de nauwkeurigheid van de uitkomsten van de survey. Het is daarom van belang om in de onderzoeksverantwoording zowel de omvang van de bruto steekproef als de netto steekproef te vermelden.

### **3.3.5 Het responspercentage en selectiviteit van de 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.

Non-respons komt altijd voor. De mate van non-respons in surveys in Nederland varieert in de praktijk tussen de 20% en 95%. Het onderwerp van een survey is een belangrijke

verklaring voor die variatie, net als de organisatie die het onderzoek uitvoert (een van de surveys met een relatief hoge respons is de Veiligheidsmonitor waarin door het CBS wordt gevraagd naar criminaliteit en veiligheidsgevoel). Een ander groot deel van de variatie in non-respons wordt verklaard door verschillen in veldwerk. Een hoge respons kan worden gehaald door: a) het inzetten van getrainde interviewers die huis-aan-huis langsgaan, b) het inzetten van beloningen, c) non-responsconversieprocedures, d) een neutrale en aantrekkelijke aanschrijfbrief, en e) veel contactpogingen op verschillende dagen/tijden en met verschillende methodes (post, telefoon, internet, huis-aan-huis).

Non-respons leidt ertoe dat sommige groepen oververtegenwoordigd zijn in de respons en andere groepen ondervertegenwoordigd. Daardoor kunnen de uitkomsten van de survey een vertekening hebben. Vertaald in praktische termen wordt de omvang van de vertekening door non-respons door twee factoren bepaald:

- **Hoeveel mensen doen niet mee?**  
Een hoger percentage non-respons leidt tot een grotere vertekening in de belangrijkste uitkomsten van het onderzoek. Met andere woorden: Is de respons bij een survey laag, dan is het risico groot dat er ernstige vertekeningen zijn in de uitkomsten. Bij een hoge non-respons kan de leefbaarheid in een provincie werkelijkheid veel beter of slechter zijn dan blijkt uit de resultaten van de survey.
- **Zijn non-respondenten anders dan respondenten?**  
Bijvoorbeeld: Uit verkiezingsonderzoek blijkt vaak dat er een sterke samenhang is tussen respons en stemgedrag. Mensen die niet meedoen aan het onderzoek gaan ook niet stemmen; mensen die wel meedoen aan het onderzoek gaan juist wel stemmen. Dit leidt ertoe dat stemmers zijn oververtegenwoordigd in de peiling, met als gevolg dat de verkiezingsopkomst te hoog wordt geschat. In de context van leefbaarheidsonderzoek in Groningen kan het zo zijn dat mensen met een laag vertrouwen in overheden/instanties veel negatiever over de leefbaarheid in de regio denken dan mensen met een hoog vertrouwen. Wanneer mensen met een laag vertrouwen in de overheid/instanties ook minder vaak meedoen aan surveys over leefbaarheid, dan leidt dit tot een vertekening in de uitkomsten van leefbaarheidsonderzoeken.

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

Een voorbeeld: stel dat slechts 40% van de mensen in de steekproef meedoet aan een verkiezingsonderzoek. Van die respondenten zegt 70% te zullen gaan stemmen. Als 40% meedoet aan het onderzoek dan is de non-respons dus 60%. De ene extreme situatie is dat al die non-respondenten niet gaan stemmen. Dan wordt het percentage stemmers in de hele steekproef:

$$0,40 \times 70\% + 0,60 \times 0\% = 28\%.$$

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

$$0,40 \times 70\% + 0,60 \times 100\% = 88\%.$$

Dus het in de respons gevonden percentage stemmers van 40% had ook 28% kunnen zijn en het had ook 88% kunnen zijn. Het werkelijk percentage stemmers zal ergens tussen 28% en 88% liggen. Een dergelijke bandbreedte maakt onderzoek in feite onbruikbaar. Gelukkig doen die extreme situaties zich in de praktijk niet vaak voor, maar het effect van non-respons kan zeker groot zijn. Tabel 1 bevat de bandbreedte voor de werkelijke percentages in de steekproef voor een reeks aan verschillende responspercentages. Duidelijk is te zien dat de bandbreedte kleiner wordt naarmate het percentage respons hoger is.

**Tabel 1. Bandbreedte werkelijke percentages in de steekproef ten gevolge van non-respons**

Gevonden	Responspercentage			
	20	40	60	80
Percentage in survey				
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

Om de non-respons te kunnen analyseren, zijn *hulpvariabelen* nodig. Dat zijn variabelen die gemeten zijn in de survey en waarvoor ook de verdeling in de doelpopulatie bekend en beschikbaar is. Hulpvariabelen zitten soms in het steekproefkader (bijvoorbeeld leeftijd en geslacht in een populatieregister, welke bekend zijn voor zowel respondenten als non-respondenten), soms kan een statistisch bureau de populatieverdeling leveren en soms kunnen de observaties van interviewers worden gebruikt (bijvoorbeeld het type woning) voor zowel respondenten als niet-respondenten.

Door de populatieverdeling van een hulpvariabele te vergelijken met de verdeling in de respons van de survey, wordt duidelijk of de respons representatief is met betrekking tot deze variabele. Als bijvoorbeeld de respons van een survey onder de hele Nederlandse bevolking voor 60% uit mannen bestaat en voor 40% uit vrouwen, dan is er iets mis. De Nederlandse bevolking bestaat immers voor 49,5% uit mannen en voor 50,5% uit vrouwen. Kennelijk hebben de mannen vaker meegedaan in het onderzoek, want ze zijn oververtegenwoordigd.

Als de respons niet representatief is met betrekking tot een aantal hulpvariabelen, dan is dit een aanwijzing dat er ook iets mis kan zijn met de representativiteit van de uitkomsten van de survey. De situatie is vooral problematisch als de doelvariabelen gecorreleerd zijn met de hulpvariabelen. Stel dat men in het zojuist genoemde onderzoek wilde weten hoeveel hamburgers een Nederlander per week eet, dan is er zeker vertekening: Mannen eten waarschijnlijk gemiddeld meer hamburgers dan vrouwen. De vertekening is waarschijnlijk veel kleiner als men wil weten hoeveel glazen melk er worden gedronken in Nederland. Er is immers waarschijnlijk geen sterk verband tussen geslacht en de hoeveelheid melk die wordt gedronken.

Een diepgaandere behandeling van de effecten van non-respons is te vinden in bijvoorbeeld [Bethlehem, Cobben, en Schouten \(2011\)](#).

### 3.3.6 Correctie voor non-respons

Als zich non-respons voordoet in een survey, dan kunnen de uitkomsten een vertekening hebben. Om daarvoor te corrigeren, moet een *weging* worden uitgevoerd. Aan elke respondent wordt een gewicht toegekend. Die gewichten worden zo berekend dat ze corrigeren voor de over- of ondervertegenwoordiging van groepen in de respons. Respondenten in oververtegenwoordigde groepen krijgen een gewicht kleiner dan 1. Respondenten in ondervertegenwoordigde groepen krijgen een gewicht groter dan 1. De resultaten van de laatstgenoemde groep tellen dus extra zwaar mee.

Een diepgaandere behandeling van de correctie voor non-respons is te vinden in bijvoorbeeld [Bethlehem et al. \(2011\)](#). In [Bijlage 4](#) van dit rapport gaan we hier ook verder op in.

### 3.3.7 Onzekerheidsmarges

De uitkomsten van een survey 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 werkelijke cijfers voor de gehele doelpopulatie. Zelfs in het ideale geval van een nette aselechte steekproef zonder non-respons is er een afwijking tussen schatting en werkelijke waarde. Dit is de 'ruis' van de steekproef.

Het is belangrijk de onzekerheidsmarges te vermelden in de onderzoeksverantwoording. Alleen dan kan onderscheid worden gemaakt tussen 'echte' effecten in de resultaten en artefacten die zijn veroorzaakt door de 'ruis' in de steekproef. Hierbij dient men te bedenken dat echte effecten beter gedetecteerd kunnen worden naarmate de omvang van de steekproef groter is. De onzekerheidsmarges zijn dan immers kleiner.

### 3.3.8 De kwaliteit van de vragenlijst

De vragenlijst is van cruciaal belang. Kleine wijzigingen in de vragen kunnen al leiden tot grote wijzigingen in de antwoordpatronen. Het is daarom belangrijk dat de vragen neutraal en begrijpelijk geformuleerd zijn en dat alle respondenten de vragen op dezelfde manier begrijpen. De vragen moeten valide zijn: Ze moeten meten wat de onderzoeker beoogt te meten.

Om de kwaliteit van de gebruikte vragenlijst te kunnen beoordelen, is het belangrijk dat een kopie van de vragenlijst in de onderzoeksverantwoording is opgenomen. Bij voorkeur is de vragenlijst vóór afname getest. Een geteste vragenlijst geeft meer vertrouwen dat deze van goede kwaliteit is. Als dit inderdaad is gebeurd, moet dit worden vermeld. Ook is het zinvol om inzicht te geven in de wijze waarop is getest.

Een van de belangrijkste aspecten van de vragenlijst is het formuleren van de tekst van de vraag. Die tekst moet begrijpelijk en te beantwoorden zijn voor iedereen, en mag de respondenten niet een bepaalde kant op sturen. Door het manipuleren van de teksten van de vragen is het mogelijk de uitkomsten van het onderzoek te beïnvloeden.

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. Zulke problemen kunnen bijvoorbeeld ontstaan als de vragenlijst is gemaakt door deskundigen die veel verstand hebben van het onderwerp van het onderzoek, maar zich niet verplaatsen in de argeloze respondent die daar veel minder in thuis is. Voorbeeld: “Bent u tevreden over de recreatieve voorzieningen in uw woonplaats?”.
- **Ondubbelzinnige vragen.**  
Een goede vraag is maar voor één uitleg vatbaar. Dubbelzinnige vragen kunnen worden vermeden door precies te vragen wat de onderzoeker wil weten. Voorbeeld: “Wanneer ging u van school af?”. Als het hier om een datum gaat, vraag dan ook naar een datum. Als het om een leeftijd gaat, vraag dan ook naar de leeftijd. En als het om een andere gebeurtenis gaat (“toen ik in verwachting was”), vraag dan ook naar die gebeurtenis.
- **Suggestieve vragen.**  
Stel geen suggestieve vragen. Respondenten moeten hun eigen antwoord kunnen geven, en niet een bepaalde kant op worden gestuurd. Daarom moeten woorden zoals ‘ook’, ‘toch’ en ‘wel’ zo veel mogelijk worden vermeden. Ook is het onjuist om in de vraagtekst deskundigen aan te halen. Respondenten zullen lang niet altijd die deskundigen durven tegenspreken. Een voorbeeld van een suggestieve vraag: “Bent u het eens met de meerderheid van de bevolking dat de gezondheidszorg in het land achteruit gaat?”.
- **Dubbele vragen.**  
Vermijd vragen waarin meer dan één vraag wordt gesteld. Voorbeeld: “Is het u bekend dat de lokale omroep een website heeft, en dat daarop een samenvatting is te vinden van al het lokale nieuws?”. Hier is sprake van twee vragen in één. De oplossing van dit probleem is simpel: splitsen in aparte vragen.
- **Vragen met een ontkenning.**  
Vermijd vragen die een ontkenning of dubbele ontkenning bevatten. Dit soort vragen zijn lastig te interpreteren. Respondenten moeten als het ware eerst een vertaalslag maken. Wat wordt er eigenlijk gevraagd? Dit soort vragen worden ook vaak verkeerd geïnterpreteerd. Voorbeeld: “Gaat u liever niet naar een café waarin niet gerookt mag worden?”.
- **Vragen die een beroep doen op de herinnering.**  
Mensen zijn niet goed in het herinneren van gebeurtenissen. Ze vergeten gebeurtenissen of plaatsen ze op een verkeerd moment in de tijd (telescoping). Dit soort vragen levert vooral problemen op als het om een relatief lange periode gaat en gevraagd wordt naar gebeurtenissen die niet zo heel belangrijk zijn en regelmatig voorkomen. Voorbeeld: “Hoe vaak heeft u in het afgelopen jaar een bezoek gebracht aan uw huisarts?”.
- **Vragen over gevoelige onderwerpen.**  
Bij het stellen van vragen over gevoelige onderwerpen loopt de onderzoeker altijd het risico dat niet het juiste antwoord wordt gegeven. Het kan zijn dat de vraag niet wordt beantwoord (ofwel item non-respons) of dat een sociaalwenselijk antwoord wordt gegeven. Bedenkt hierbij dat gevoelige vragen beter worden beantwoord als de respondenten helemaal op zichzelf de vragenlijst invullen (bijvoorbeeld vragenlijst per post of online vragenlijst) en er geen interviewers aanwezig zijn. Voorbeeld: “Bezoekt u weleens pornowebsites op het internet?”.

- **Hypothetische vragen.**  
In sommige gevallen zijn onderzoekers geïnteresseerd in de effecten van veranderingen in beleid. Stel dat de overheid de accijns op alcohol wil verdubbelen, dan kan een survey worden gedaan om de reactie van respondenten daarop te peilen. Dikwijls worden dan vragen gesteld als: “Stel dat de prijs van een bier in een café met 70% zou worden verhoogd, zou u dan minder gaan drinken?” Dit is een hypothetische vraag die erg moeilijk te beantwoorden is voor respondenten. Enerzijds is de vraag moeilijk omdat de respondent wellicht nooit heeft nagedacht over dit onderwerp, anderzijds is het moeilijk om een houding of gedrag in de toekomst te voorspellen. Het is beter dit soort hypothetische vragen te vermijden.

Een vragenlijst bestaat gewoonlijk uit een hele reeks vragen. Het is dan belangrijk in welke volgorde de vragen worden gesteld. Het antwoord op een vraag kan namelijk worden beïnvloed door eerder gestelde vragen. Als het bijvoorbeeld in eerder gestelde vragen alleen maar gaat over de negatieve gevolgen van aardbevingen, dan kan de respondent steeds negatiever gaan denken over zijn/haar algemene tevredenheid met de woonomgeving. De respondent wordt negatief *geframed*. Als een onbevooroordeeld antwoord op een vraag belangrijk is, dan is het daarom beter die vraag vooraan in de vragenlijst te stellen. Ook is het goed te beginnen met algemene vragen en later in het onderzoek de meer specifieke vragen te stellen.

Voor een objectiever beeld is het dus beter om zowel aan negatieve als positieve aspecten aandacht te besteden in de vragenlijst. Dat geldt ook voor de vragen zelf. Als een gesloten vraag een heleboel negatieve antwoordopties heeft, en maar weinige positieve, dan valt het totaalbeeld vaak negatief uit. Het is beter om een meer gebalanceerde reeks antwoordopties aan te bieden.

Veel vragenlijsten worden tegenwoordig via het internet aangeboden. Dat betekent dat er geen interviewers aanwezig zijn om personen in de steekproef te overtuigen om mee te doen. Ze kunnen ook geen hulp bieden bij het invullen van de vragenlijst. De respondenten moeten het helemaal in hun eentje doen. Daarbij kunnen allerlei verschijnselen leiden tot fouten in de antwoorden op de vragen (*meetfouten* worden die wel genoemd). Krug (2006) beschrijft hoe mensen websites lezen. Dat is uiteraard ook van toepassing op online surveys:

- Respondenten hebben weinig of geen belangstelling voor het onderwerp van de survey.
- Meedoen aan het onderzoek is niet echt belangrijk voor hen.
- Ze lezen de vragen niet nauwkeurig, maar scannen ze slechts globaal.
- Ze weten dat er geen straf staat op het geven van onjuiste antwoorden.
- Ze zoeken niet uit hoe de vragenlijst precies werkt, maar modderen voort van vraag naar vraag en proberen zo het eind te halen.

Dit alles leidt ertoe dat respondenten niet het beste antwoord geven (het antwoord dat zo dicht mogelijk bij de werkelijkheid ligt), maar dat ze met een minimale inspanning het eerste redelijke antwoord kiezen dat ze acceptabel vinden. Dit wordt wel *satisficing* genoemd. *Satisficing* komt voor in allerlei vormen:

- **Primacy effect:**  
De respondent kiest in een lange lijst van mogelijke antwoorden een antwoord vooraan in de lijst.

- **Acquiescence:**  
De respondent is geneigd in te stemmen met een stelling in een opinievraag ongeacht de inhoud ervan.
- **Endorsing the status quo:**  
Als een respondent wordt gevraagd zijn/haar mening te geven over een mogelijk verandering, zal hij/zij meer geneigd zijn te kiezen voor de optie 'geen verandering'.
- **Preference for the middle option:**  
Als een respondent naar zijn/haar mening wordt gevraagd, zal hij/zij geneigd zijn te vluchten naar de middelste, neutrale optie die aangeeft dat hij/zij geen duidelijke mening heeft.
- **Straightlining:**  
Als een reeks vragen met allemaal dezelfde antwoordmogelijkheden wordt aangeboden in de vorm van een matrix of grid, dan is de respondent geneigd alle antwoorden in dezelfde kolom te kiezen. Het moeten beantwoorden van veel matrixvragen is meestal niet bevorderlijk voor de motivatie van de respondenten.
- **Don't know:**  
Als bij een opinievraag ook de optie 'weet niet' wordt aangeboden, zullen veel respondenten daar uit gemakzucht voor kiezen. Het is een makkelijke uitweg. Het weglaten van deze optie is ook vaak niet goed, omdat er nu eenmaal respondenten zijn die het echt niet weten en die mogen niet worden gedwongen tot een antwoord dat niet het hunne is.
- **Arbitrary answers:**  
Vooral bij vragen waarbij alle in aanmerking komende antwoorden moeten worden aangekruist (meerdere antwoorden mogelijk), beperkt de respondent zich uit gemakzucht tot maar een paar mogelijkheden (en dan vaak ook nog tot de mogelijkheden vooraan in de lijst).

Al deze effecten kunnen leiden tot onjuiste en minder juiste antwoorden. Dat heeft een negatief effect op de kwaliteit van de antwoorden en dus op de kwaliteit van de uitkomsten van de survey. In algemene zin is dit gedrag alleen te voorkomen door goede vragen te stellen. Ook helpt het om repetitieve vragen (bijvoorbeeld stellingen met telkens dezelfde antwoordcategorieën) te vermijden.

Er moet wel worden opgemerkt dat er ook aspecten zijn waarop online surveys het beter doet dan surveys met interviewers. Een voorbeeld daarvan is gevoelige vragen. Sociaalwenselijke antwoorden worden eerder gegeven op gevoelige vragen als een interviewer de vragen persoonlijk stelt. Bij een online survey is de respondent geneigd om eerlijker te antwoorden. Meer over de methodologische haken en ogen van online surveys kan worden gevonden in [Bethlehem en Biffignandi \(2012\)](#) en in [Couper \(2008\)](#).

In onze review van de vijf studies zullen we niet iedere vraag in de vragenlijst in detail beoordelen op alle bovengenoemde criteria. We zullen meer in algemene zin wel beoordelen of er veel of weinig problematische vragen worden gesteld. Ook zullen we kijken naar de vragenlijst als geheel. Is de volgorde van de vragen logisch, is er een goede verhouding tussen de verschillende typen vragen die zijn gesteld, en vormen de vragen die worden gesteld een goede basis voor het geven van een antwoord op de door de onderzoekers gestelde onderzoeksvragen?

### 3.3.9 De analyse van de data

Op basis van data kunnen namelijk allerlei analyses gedaan worden (bijvoorbeeld visualisaties, en het onderzoeken van relaties of verbanden door de tijd heen). Welke analyse wordt gedaan, hangt onder andere af van onderzoeksvragen die centraal staan in het onderzoek, en van het type vragen in de survey en de bijbehorende antwoordmogelijkheden. We zullen bij elk onderzoek dat we reviewen, kort bespreken of de analyses die worden uitgevoerd van goede kwaliteit zijn. We zullen niet inhoudelijk kijken of de analyses zinvol zijn of niet, maar ons beperken tot de vraag of er gegeven de data die uit het survey-onderzoek voortkomen, juiste keuzes worden gemaakt in de analyses. Ook bekijken we of er wellicht aanvullende analyses mogelijk zijn met de data die meer inzicht kunnen geven in het effect van aardbevingen op de leefbaarheid in Groningen.

## 3.4 Relevantie voor toekomstige effectmetingen naar leefbaarheid

De tweede onderzoeksvraag in dit onderzoek is: “Op welke manier kan toekomstig onderzoek tegemoetkomen aan geconstateerde zwaktes en moeilijkheden van bestaand onderzoek? Aanbevelingen die hieruit voortvloeien moeten erop gericht zijn om in de toekomst goede effectmetingen te kunnen doen naar veranderingen in de leefbaarheid in het aardbevingsgebied, rekening houdend met veranderend beleid en andere maatschappelijke veranderingen (bijvoorbeeld bevolkingskrimp).”

Goed opgezette en uitgevoerde onderzoeken zullen over het algemeen meer ‘best practices’ voor toekomstig leefbaarheidsonderzoek kunnen opleveren. Aan de hand van kanttekeningen bij gereviewde onderzoeken kunnen we ook tot aanbevelingen komen over de methodologische opzet van toekomstig onderzoek. Toch is het niet zo dat de conclusies die we met behulp van de methodologische meetlat over de gereviewde onderzoeken trekken, een op een te vertalen zijn naar aanbevelingen voor de toekomst.

De vijf studies die in dit rapport worden gereviewd zijn niet alleen zijn geselecteerd op basis van het feit dat zij leefbaarheid proberen te meten. Een van onze selectiecriteria was juist dat een verscheidenheid aan andere onderwerpen in de surveys aan bod komt. De surveys die we in het volgend hoofdstuk uitgebreid zullen reviewen, richten zich bijvoorbeeld ook op de (psychische) gezondheid van inwoners (Gronings Perspectief), tevredenheid met de shadeafhandeling (APE) en dynamiek op de woningmarkt (OTB Delft). Hoewel deze andere onderwerpen te maken hebben met het onderwerp leefbaarheid zijn de onderzoeken niet per se opgezet om de leefbaarheid in Groningen in kaart te brengen. Hierdoor zijn deze onderzoeken niet zonder meer bruikbaar voor toekomstig leefbaarheidsonderzoek. Het kan in theorie dus zo zijn dat we concluderen dat we de kwaliteit van een onderzoek (opzet, uitvoering, rapportage) heel goed vinden gegeven het doel van het onderzoek, en de resultaten relevant voor beleid en investeringen van overheden, maar ook vinden dat de bruikbaarheid voor toekomstig leefbaarheidsonderzoek en het evalueren van de effecten van op leefbaarheid gericht beleid beperkt is. Het belangrijkste doel van het onderzoek uitgevoerd door APE was bijvoorbeeld om de tevredenheid met shadeafhandeling onder mensen die een schademelding hadden gedaan bij de NAM/het Centrum Veilig Wonen in kaart te brengen. Zelfs als we bij dit onderzoek tot de conclusie zouden komen dat het onderzoek uitmuntend is uitgevoerd, en er geen enkele methodologische kritiek te leveren zou zijn,

dan nog is de methodologie van dit onderzoek niet zomaar toepasbaar op of relevant voor toekomstig onderzoek naar de leefbaarheid.

Het meten van de effecten van beleidsmaatregelen of investeringen is verder ingewikkeld. Het woord effect impliceert dat een bepaalde oorzaak (een beleidsmaatregel of investering) al dan niet een causaal gevolg heeft op leefbaarheid. Voor het doen van zulke causale uitspraken moet aan drie voorwaarden voldaan zijn:

- Er moet een verband zijn tussen twee of meer constructen. Bijvoorbeeld: Een investering in sociale voorzieningen leidt tot een toegenomen waargenomen leefbaarheid van de woonomgeving.
- De oorzaak (een investering) moeten voorafgaan aan het gevolg (de leefbaarheid van de woonomgeving).
- Er mogen geen andere oorzaken zijn die het verband tussen twee of meer constructen kunnen verklaren. Bijvoorbeeld: Economische veranderingen moeten als mogelijke oorzaak van de hogere leefbaarheid van de woonomgeving worden uitgesloten.

Aan de eerste voorwaarde kan worden voldaan door in een gebied waar bijvoorbeeld investeringen zijn gedaan te meten of die investeringen gepaard gaan met een hogere leefbaarheidsscore. Het is wel noodzakelijk dat de doel- of uitkomstvariabele wordt gespecificeerd: Is het belangrijk dat effecten worden gemeten en gevonden op veiligheidsgevoel, tevredenheid met de woonomgeving, gezondheid, of nog meer of andere constructen? Soms is het van tevoren moeilijk vast te stellen welke effecten precies worden verwacht. Toch is het definiëren van doel- of uitkomstvariabelen een essentiële stap voor het mogelijk maken van effectonderzoek.

Om aan de tweede voorwaarde te voldoen is het belangrijk dat er effectmetingen worden gedaan door de tijd heen. Metingen moeten eigenlijk al beginnen voordat de beleidsmaatregel of investering wordt ingezet, en op meerdere momenten bij dezelfde groep mensen afgenomen worden om veranderingen te meten. Longitudinaal onderzoek, waar dezelfde mensen door de tijd heen worden gevolgd is daar geschikt voor.

Aan de derde voorwaarde voor causaliteit is het lastigst te voldoen als er geen experimenteel onderzoeksdesign wordt gebruikt. Hoe kunnen alle mogelijke andere oorzaken voor veranderingen worden uitgesloten? Er zijn allerlei mogelijke redenen te bedenken waarom door de tijd heen de leefbaarheid kan verbeteren of verslechteren, zeker op lokaal niveau. Het is heel moeilijk om alle mogelijke alternatieve verklaringen uit te sluiten. Toch zijn er manieren waarop veel mogelijke oorzaken kunnen worden uitgesloten. De eerste manier is door mogelijke alternatieve verklaringen zelf te meten. Zo kan er in de vragenlijst worden gevraagd naar mogelijke alternatieve verklaringen (bijvoorbeeld: Zijn er lokale bewonersinitiatieven die zijn gestart op de leefbaarheid te verbeteren), of kan er op geaggregeerd niveau worden gekeken naar alternatieve verklaringen (bijvoorbeeld: Zijn er winkels bijgekomen of verdwenen, wat is de groei of daling van inkomens in een gebied?). De tweede manier is door een goede vergelijkingsgroep te zoeken. Die vergelijkingsgroep kan op geografisch kleine schaal bestaan. Dorp A en B in het aardbevingsgebied kunnen bijvoorbeeld vergeleken worden, waarbij investeringen vooral zijn gedaan in dorp A. Vergelijkingen kunnen ook op regionaal niveau plaatsvinden (bijvoorbeeld: Het aardbevingsgebied en het niet-aardbevingsgebied in Groningen, krimp en niet-krimpgebieden, of de

Noordelijke provincies). Welke vergelijkingsgroep relevant is, hangt af van het soort effecten dat men beoogt te meten en van de specifieke beleidsmaatregel of investering waarvoor is gekozen. Om alternatieve verklaringen zoveel mogelijk uit te sluiten kan het beste gebruik worden gemaakt van een combinatie van strategieën: Alternatieve verklaringen moeten zoveel mogelijk worden gemeten, en er moet geprobeerd worden een zo goed mogelijke vergelijkingsgroep te vinden—een vergelijkingsgroep waarbij de beleidsmaatregel of investering niet is uitgevoerd.

Om onze tweede onderzoeksvraag goed te kunnen beantwoorden, zullen we gebruik maken van de sterke en minder sterke kanten van ieder onderzoek. Ook zullen we expliciet bespreken in hoeverre (onderdelen van) onderzoeken gebruikt zouden kunnen worden in toekomstige effectmetingen. In Hoofdstukken 5 en 6 geven we tenslotte verder antwoord op onze twee centrale onderzoeksvragen.

## 4.

## Onderzoeksresultaten

Het grootste deel van dit hoofdstuk zal bestaan uit afzonderlijke besprekingen van de vijf gereviewde survey-onderzoeken. We hebben de vijf geselecteerde onderzoeken langs de methodologische meetlat uit [sectie 3.3](#) en de criteria uit [sectie 3.4](#) gelegd en gaan in [secties 4.1-4.5](#) van dit hoofdstuk per onderzoek in op onze bevindingen. De volgorde waarin we de onderzoeken bespreken is willekeurig, het betreft dus geen rangordening waarbij het eerste onderzoek dat we noemen het ‘beste’ is, of vice versa. In [secties 4.1-4.5](#) geven we voor elk gereviewd onderzoek eerst wat beschrijvende informatie over het desbetreffende onderzoek. Hierop volgt een samenvatting van onze review en onze bevindingen. We bespreken vervolgens de sterke en zwakke punten van het onderzoek in meer detail, daarbij gebruikmakend van de methodologische meetlat. We houden hierbij grotendeels de volgorde van de meetlat uit [sectie 3.3](#) aan, maar kijken daar soms vanaf als dit de leesbaarheid van de review ten goede komt (indien keuzes in de onderzoeksopzet bijvoorbeeld consequenties hebben voor verdere punten op de meetlat). We hebben er ook voor gekozen om twee punten uit de meetlat—de opdrachtgever/financier en de onderzoeksverantwoording—niet per onderzoek afzonderlijk te bespreken, maar deze geïntegreerd voor de vijf onderzoeken te bespreken (in onderstaande paragrafen voorafgaand aan [sectie 4.1](#)). Tenslotte bespreken we per onderzoek in hoeverre het onderzoek bruikbaar is voor het opzetten van toekomstige effectmetingen naar leefbaarheid (zie ook [sectie 3.4](#)). Voor uitleg van begrippen verwijzen we naar [Hoofdstuk 3](#).

### Opdrachtgever/financier van de vijf onderzoeken

Van de vijf onderzoeken die we hebben gereviewd is één onderzoek landelijk: dit betreft de ‘Leefbaarometer 2.0’. Dit is een onderzoek waarbij allerlei landelijke indicatoren, afkomstig uit registers, en het WoON-survey samen worden beschouwd. Dit onderzoek wordt periodiek uitgevoerd door onderzoeksbureau RIGO en de Atlas voor gemeenten. Financier en opdrachtgever is het Ministerie van Volkshuisvesting Ruimtelijke Ordening en Milieu. De Leefbaarometer is niet opgezet met het doel de effecten van aardbevingen te meten, of de leefbaarheid in Groningen. Het betreft juist een algemeen, landelijk onderzoek.

De andere vier onderzoeken die we hebben gereviewd zijn in meer of mindere mate opgezet naar aanleiding van de aardbevingsproblematiek in Groningen. De Dialoogtafel Groningen is opdrachtgever geweest van twee onderzoeken. Het eerste van deze onderzoeken betreft: ‘Wonen en aardbevingen in Groningen’, dat is uitgevoerd door OTB (onderdeel van TU Delft) en onderzoeksbureau CMO STAMM. Daarnaast heeft de Dialoogtafel het onderzoek naar de ‘Tevredenheid van de schadeafhandeling door de NAM’, uitgevoerd door onderzoeksbureau APE gefinancierd. ‘Gronings Perspectief’ is uitgevoerd door de Rijksuniversiteit Groningen (RUG) in samenwerking met de GGD en Onderzoeksbureau Onderzoek en Statistiek Groningen in opdracht van de Nationaal Coördinator Groningen (NCG). Bij deze drie onderzoeken is ervoor gekozen om onafhankelijke stuurgroepen in te stellen die het onderzoek begeleiden. Deze stuurgroepen bestaan meestal uit belanghebbende bewoners van het aardbevingsgebied, bestuurders uit de provincie Groningen, wetenschappers, en ambtenaren van de semioverheid. De aanwezigheid van een stuurgroep geeft meer zekerheid dat de onderzoeken onafhankelijk zijn uitgevoerd.

Tenslotte is het onderzoek 'Leefbaarheid en bevolkingskrimp in Groningen' uitgevoerd door het Sociaal Planbureau Groningen (SPG). Het SPG vormt een onderdeel van onderzoeksbureau CMO STAMM, dat ook betrokken is geweest bij het onderzoek 'Wonen en aardbevingen in Groningen'. Het Planbureau heeft een onafhankelijke positie als het gaat om de vraagstelling en opzet van de onderzoeken die worden uitgevoerd. Het Planbureau wordt gefinancierd door de provincie Groningen.

Al met al kan worden gesteld dat de opdrachtgevers en financiers van de onderzoeken die wij hebben gereviewd allemaal onderdeel uitmaken van de (semi-) overheid. In het geval van de Dialoogtafel en NCG betreffen het organisaties die in het leven zijn geroepen naar aanleiding van de aardbevingsproblematiek. De Dialoogtafel had als doel overleg te faciliteren tussen alle betrokken partijen binnen het aardbevingsdossier in Groningen en kan in die zin worden gezien als neutraal. De doelstelling van het NCG sluit hier nauw bij aan. Het Sociaal Planbureau Groningen heeft eveneens een onafhankelijke positie als het gaat om het opzetten van onderzoek, en is alleen indirect afhankelijk van de overheid voor financiering.

In algemene zin is het niet zo dat onderzoekers geen enkel belang hebben bij onderzoeken. Het kunnen opzetten en uitvoeren van nieuw onderzoek op zich is al een belang van onderzoekers, en als burger kunnen onderzoekers ook een duidelijke mening hebben over het aardbevingsdossier. Hoewel alle onderzoekers als burger een mening zullen hebben over het aardbevingsdossier, voorzien de regels van (wetenschappelijk) onderzoek erin dat onderzoekers hun onderzoek transparant en onafhankelijk uitvoeren.

### **Onderzoeksverantwoording van de vijf onderzoeken**

Bij de vijf onderzoeken die wij hebben gereviewd bestaat er in alle gevallen een goede onderzoeksverantwoording. In de meeste gevallen bestaat deze uit een onderzoeksrapport dat (online) is gepubliceerd, soms voorzien van bijlages met bronmateriaal (zoals de afgenomen vragenlijst). In [Hoofdstuk 3](#) hebben we onze werkwijze voor dit rapport uitgelegd: Het grootste deel de onderzoeken was te evalueren aan de hand van deze publiek beschikbare informatie. In het geval van het gereviewde onderzoek van het Sociaal Planbureau Groningen was de publiek beschikbare informatie summier en onvoldoende om de gevolgde methodologie goed te beoordelen. Bij alle onderzoeken hebben we na inventarisatie van de ontbrekende informatie om aanvullende informatie gevraagd, en in alle gevallen gekregen. In het vervolg van dit hoofdstuk bespreken we elk van de vijf onderzoeken die door ons gereviewd zijn.

## **4.1 Gronings Perspectief (RUG)**

Gronings Perspectief is opgezet in 2015 en wordt gecoördineerd door de Rijksuniversiteit Groningen in samenwerking met de GGD Groningen en Onderzoeksbureau Onderzoek en Statistiek Groningen. Gronings Perspectief kent meerdere onderzoekslijnen, waarbij we ons in de huidige review richten op het panelonderzoek binnen Gronings Perspectief. In dit panelonderzoek worden panelleden vijf maal ondervraagd over uiteenlopende thema's. Hoewel er vragen worden gesteld over leefbaarheid, ligt de nadruk in het onderzoek op het in kaart brengen van de psychosociale en gezondheidsgevolgen van gaswinning voor de inwoners van Groningen. Op het moment van schrijven van dit rapport (juli 2017) loopt het onderzoek nog en zijn er drie metingen afgerond. In

dit rapport reviewen we de eerste drie panelmetingen van Gronings Perspectief en bijbehorende rapportages, in de context van de hele studie. Aan het einde van 2017 zullen alle vijf panelmetingen zijn uitgevoerd. In [Bijlage 6](#) is een reactie opgenomen van de onderzoekers van Gronings Perspectief op de nu volgende bevindingen.

### **Algemeen oordeel**

Omdat een deel van de dataverzameling van Gronings Perspectief nog in 2017 wordt uitgevoerd, kunnen we het Gronings Perspectief onderzoek niet in zijn geheel evalueren. Wel kunnen we concluderen dat de onderzoeksopzet van Gronings Perspectief van methodologisch zeer hoog niveau is. De doelpopulatie is goed gekozen; de studie wordt longitudinaal uitgevoerd, bestrijkt naast leefbaarheid een breed pallet aan onderwerpen, en er is een expliciete koppeling met het langer lopende Lifelines onderzoek. De kwaliteit van enkele individuele vragen zou wel verbeterd kunnen worden, en de gehanteerde weegmethode kan worden geoptimaliseerd. Dit betreffen twee kleine kanttekeningen. Een grotere kanttekening willen we plaatsen bij de hoge non-respons. De respons in Gronings Perspectief is vergelijkbaar met de andere studies die we gereviewd hebben. Toch zien we bij een respons van rond de 20% een risico op vertekening, die moeilijk of niet te corrigeren valt door te wegen op sociaal-demografische achtergrondvariabelen. Dit onderzoek zou zeer bruikbaar zijn als blauwdruk voor toekomstig onderzoek naar effecten van beleid op leefbaarheid.

### **Doelpopulatie**

De doelpopulatie bestond uit inwoners van de 23 Groningse gemeenten van 16 jaar en ouder. Gronings Perspectief is erop gericht een vergelijking te maken tussen a) gebieden waar panden meer en minder schade hebben en b) mensen die meer of minder schade hebben. In de analyses wordt er derhalve meestal gegeneraliseerd naar deze groepen.

### **Steekproeftrekking**

Voor dit onderzoek is een aselechte steekproef uit de doelpopulatie getrokken, via de Basisregistratie Personen (BRP). De BRP betreft een register waar alle in Nederland wonende personen staan ingeschreven. In de praktijk kunnen er soms foute registraties in BRP staan, maar die zijn zeldzaam. De BRP is een van de meest complete registers die in Nederland beschikbaar is.

In totaal zijn 25000 bewoners in de Provincie Groningen per brief persoonlijk uitgenodigd om lid te worden van het bewonerspanel. Er zijn relatief meer uitnodigingen verstuurd naar inwoners in dunbevolkte postcodegebieden (d.w.z. stratificatie), om ervoor te zorgen dat er voldoende respondenten uit deze gebieden aan het onderzoek zouden deelnemen. Alle panelleden ontvingen vervolgens bij elke meting via e-mail een uitnodiging om een online vragenlijst te beantwoorden. Personen zonder internet of met traag internet konden op aanvraag een schriftelijke vragenlijst toegestuurd krijgen (ongeveer 5% koos hiervoor). Er waren geen beloningen voor respondenten.

Bewoners die geen uitnodiging kregen om lid te worden van het bewonerspanel, maar wel wilden meedoen, konden zich op een website aanmelden. Deze 127 zelfaanmelders zijn in de data-analyses van de resultaten niet meegenomen, omdat ze op een aantal punten verschilden van de aselechte steekproef.

## Resultaat veldwerk

### *Panel*

- Bruto steekproef: 25000 personen zijn uitgenodigd voor het panel.
- Netto steekproef (respons): 4556 personen hebben de uitnodiging geaccepteerd.
- Percentage respons: 18,2%.
- Ouderen zijn oververtegenwoordigd en jongeren zijn ondervertegenwoordigd. Hoger opgeleiden zijn oververtegenwoordigd.

### *De 3 metingen binnen het panel:*

Panelleden kregen in de drie metingen een uitnodigingsbrief met een URL naar de vragenlijst, en daarnaast nog twee herinneringen. De respons per meting is hieronder weergegeven:

- **Meting 1 (februari-april 2016):**  
3927 panelleden hebben de vragenlijst ingevuld (percentage respons conditioneel op deelname aan onderzoek: 86%; respons uitgaande van de bruto steekproef is 15,7%).
- **Meting 2 (juni-juli 2016):**  
3156 panelleden hebben de vragenlijst ingevuld (percentage respons 69% conditioneel op deelname aan onderzoek; respons uitgaande van de bruto steekproef is 12,6%).
- **Meting 3 (oktober-november 2016):**  
2638 panelleden hebben de vragenlijst ingevuld (percentage respons 58% conditioneel op deelname aan onderzoek; respons uitgaande van de bruto steekproef is 10,6%).

In totaal hebben 2310 panelleden (percentage response conditioneel op deelname aan het onderzoek: 50%; 9,2% uitgaande van de bruto steekproef van 25000) alle drie de vragenlijsten ingevuld. In metingen 2 en 3 is er uitval uit het panel (meer dan 10% per wave). Er is onderzoek gedaan naar de selectiviteit door uitval en met name jongeren vallen uit.

## Correctie

Als gevolg van het gebruik van ongelijke selectiekansen en de non-respons van ongeveer 85% is er grote kans op vertekening in het onderzoek.

Volgens het tweede wetenschappelijke rapport is gecorrigeerd voor de ondervertegenwoordiging van jongere deelnemers en de oververtegenwoordiging van hoger opgeleide deelnemers. Dit is gedaan door de variabelen leeftijd en opleidingsniveau als covariaten (d.w.z. extra verklarende factoren) mee te nemen in analyses. Hiernaast is er om consequenties van uitval in het panel te adresseren gewogen voor leeftijd en geslacht, en hoeveelheid schade per viercijferig postcodegebied.

In andere analyses is er gekozen voor wegingen. Er is gekozen voor weging door middel van poststratificatie. Hierin wordt de steekproef ingedeeld in categorieën op basis van een aantal variabelen. Er is op deze manier gecorrigeerd voor drie variabelen: 1) adressendichtheid op postcode (volledige postcode, 4 cijfers en twee letters) aan de hand van gegevens van het CBS, 2) geslacht, en 3) leeftijdscategorieën. Tenslotte is er ook gewogen op mate van schade. Hiervoor is het schaderegister van de NAM/het Centrum

Veilig Wonen gebruikt, waarin schademeldingen worden bijgehouden. In Gronings Perspectief lijken mensen met meervoudige schade enigszins oververtegenwoordigd te zijn ten opzichte van dit register. De onderzoekers hebben aangegeven dat uit analyses met behulp van de landelijke GGD-monitor, waarin deels dezelfde vragen zijn gesteld als in Gronings Perspectief, dezelfde uitkomsten komen. Dat wil zeggen dat ook daar meervoudige schades oververtegenwoordigd zijn ten opzichte van het schaderegister. Omdat het schaderegister ook een belangrijke rol speelt in het onderzoek van APE naar de tevredenheid met de schadeafhandeling, gaan we dieper in op de kwaliteit van het register en de rol daarvan in non-responscorrecties in [sectie 4.5](#).

Er zijn verschillende manieren om te corrigeren voor ongelijke selectiekansen en de non-respons door middel van wegingen. Het is gebruikelijk om in steekproeven met verschillende selectiekansen voor respondenten eerst gewichten te berekenen die corrigeren voor die verschillen, en daarna non-responsgewichten. In Gronings Perspectief wordt meteen gebruik gemaakt van poststratificatie, waarbij wordt gekeken naar adressendichtheid. Bij poststratificatie worden groepen die door problemen in het veldwerk ondervertegenwoordigd zijn geraakt, zwaarder meegeteld in de analyse. Als er op meerdere variabelen tegelijk gewogen wordt, betekent dat dat er al snel veel categorieën respondenten ontstaan (bijvoorbeeld een groep jonge mannen wonend in de grote stad en een groep jonge mannen wonend in een dorp). Er is soms geen informatie over de verdeling in de populatie voor al deze, heel specifieke categorieën, wat het maken van specifieke gewichten voor die categorieën moeilijk maakt. Er moet dan worden gekozen voor te brede categorieën. Bovendien kan het zijn dat bepaalde groepen erg weinig voorkomen in de data, wat kan resulteren in grote variatie tussen de gewichten. Dit laatste lijkt in Gronings Perspectief mee te vallen.

In gevallen waarin gecorrigeerd wordt voor veel variabelen tegelijkertijd is het beter gebruik te maken van multiplicatief wegen. Dit is een stapsgewijze procedure, waarbij in het weegmodel in stappen wordt geprobeerd de data representatief te maken voor een heel aantal variabelen. Meer informatie over de technische uitvoering van verschillende weegmethodes en bijbehorende voor- en nadelen is opgenomen in [Bijlage 4](#) van dit rapport.

Het is belangrijk aan te tekenen dat een verandering van weegmodel slechts heel zelden leidt tot een inhoudelijk andere uitkomst van een onderzoek. Belangrijker dan het model zijn de variabelen waarop wordt gewogen. In Gronings Perspectief is op een beperkte set variabelen gewogen. Leeftijd, geslacht, adressendichtheid en de mate van schade alleen verklaren waarschijnlijk niet alle verschillen die er zijn tussen respondenten en non-respondenten. Ook zijn deze variabelen waarschijnlijk niet sterk gecorreleerd met de centrale uitkomsten van Gronings Perspectief (zoals psychosociale gezondheid) of leefbaarheid. Goede extra weegvariabelen zouden betrekking kunnen hebben op de kenmerken van de woning. Ook zou er expliciet kunnen worden gecorrigeerd voor de specifieke longitudinale non-respons door variabelen uit eerdere metingen te gebruiken in een extra correctieprocedure. Al met al zou de weegmethode die wordt gebruikt in dit onderzoek geoptimaliseerd kunnen worden.

## Vragenlijst

De vragenlijsten zijn opgenomen in de onderzoeksverantwoording van Gronings Perspectief. Dit betreft de vragenlijsten van de eerste, tweede, en derde meting.

De uitnodigingsbrief ontbreekt in de rapportage, maar is na een verzoek door de onderzoekers opgestuurd. De uitnodiging voor deelname aan het panel en de uitnodigingen voor de verschillende metingen zijn neutraal geformuleerd. Opvallend is dat, hoewel er gebruik wordt gemaakt van een panelonderzoek waarbij dezelfde mensen meerdere keren ondervraagd worden, de meeste vragen niet herhaald worden in de verschillende metingen. Wij baseren ons hier alleen op de eerste drie metingen van Gronings Perspectief; de onderzoekers hebben aangegeven dat in metingen 4 en 5 meer herhaling van vragen zal voorkomen ten behoeve van longitudinale analyses. Ook geven zij aan dat het niet het enige doel is van Gronings Perspectief om longitudinale metingen te doen. Er is expliciet ook ruimte voor het opnemen van (verdiepende) vragen die door belanghebbenden, zoals de GGD, stuurgroep, en bewonersgrepen als relevant worden gezien. Dit proces van co-creatie is kenmerkend voor Gronings Perspectief.

In de eerste vragenlijst in Gronings Perspectief wordt vooral gebruik gemaakt van vragen die al eerder afgenomen zijn in andere onderzoeken (bijvoorbeeld in onderzoek door het CBS of in eerder wetenschappelijk onderzoek). Vragen zijn dus vaak al gevalideerd, en daarmee zijn de uitkomsten uit Gronings Perspectief vaak vergelijkbaar met die van andere onderzoeken. Voorbeelden hiervan zijn de vragen over de ervaren gezondheid. De volgorde van de onderwerpen en vragen is vaak logisch in deze en andere metingen van Gronings Perspectief. Er wordt begonnen met algemene vragen over leefbaarheid en ervaren gezondheid, en specifieke vragen over aardbevingen volgen aan het eind. Een volgorde van algemene vragen naar specifieke vragen werkt beter dan andersom. Wanneer begonnen wordt met vragen over aardbevingen, kunnen de verdere algemene vragen over leefbaarheid en gezondheid immers gekleurd worden door de antwoorden van de respondent op de vragen over aardbevingen.

In de tweede vragenlijst zijn veel nieuwe vragen opgenomen, die veelal niet uit eerder onderzoek komen en speciaal voor het onderzoek zijn opgesteld. Deze vragen gaan voornamelijk over mogelijkheden om de leefbaarheid in Groningen te verbeteren. Dit zijn vaak hypothetische vragen (“Stel dat de overheid de volgende regeling (...) wil doorvoeren. In hoeverre vindt u dat een goed idee?” of “Waardoor zou u zich veiliger gaan voelen”). Dit zijn voor de respondent heel lastige vragen, en de antwoorden op zulke vragen zijn weinig zinvol als het erom gaat beleidsvoorkeuren van respondenten te schatten.

In de derde vragenlijst worden meer vragen uit de eerste twee metingen herhaald, en wordt een aantal vragen gesteld over de toekomst en het welzijn van Groningers.

In het algemeen zijn de vragen die gesteld zijn in Gronings Perspectief neutraal geformuleerd, is er goed nagedacht over de volgorde van vragen, en zijn veel vragen afkomstig uit ander onderzoek, wat vergelijkingen mogelijk maakt. Er wordt veel gebruik gemaakt van stellingen (bijvoorbeeld “In hoeverre waren onderstaande personen of groepen in staat u op te beuren (...)” met als antwoordopties een 5-puntsschaal die loopt van [helemaal niet-heel erg]. Hoewel dit soort stellingen veel in sociaalwetenschappelijk onderzoek gebruikt wordt, en de stellingen geschikt zijn om de wensen, meningen en gevoelens van respondenten in kaart te brengen, zou er bij het meten van andere constructen in Gronings Perspectief een meerwaarde zijn om naast de huidige vragen in Gronings Perspectief, ook meer feitelijke indicatoren op te nemen. Als het bijvoorbeeld gaat om vragen over de gezondheid van respondenten wordt er door middel van ‘vague

quantifiers' [nooit, zelden, soms, vaak, meestal, voortdurend] gevraagd hoe vaak iemand last heeft gehad van een reeks gezondheidsklachten. In aanvulling op deze vragen kan het zinvol zijn ook te vragen naar de exacte frequentie van dit soort klachten te vragen. Deze extra vragen zullen niet direct betere informatie opleveren, maar wel aanvullende inzichten kunnen opleveren (bijvoorbeeld via koppeling aan registerdata).

## Analyse

Er wordt in het eerste wetenschappelijke rapport van Gronings Perspectief gezegd dat wegingen geen effect hebben op de meeste resultaten, en dat de wegingen daarom achterwege zijn gebleven in de rapportage. In een aantal andere analyses wordt wel gewogen (zie de eerdere correcties sectie). Onduidelijk is waarom er soms wel en soms niet gewogen wordt in de analyses.

De onderzoekers hebben gebruik gemaakt van geavanceerde statistische technieken (SEM, multilevel-analyses), die correct uitgevoerd zijn en dit onderzoek een duidelijke meerwaarde geven, omdat deze analyses inzicht geven in complexe verbanden tussen variabelen.

Gronings Perspectief is uniek in de zin dat respondenten op vijf verschillende momenten worden ondervraagd. Dit biedt mogelijkheden om te onderzoeken of de leefbaarheid verandert, en of die verandering wellicht anders is voor specifieke groepen respondenten. De onderzoekers hebben aangegeven dat de meeste longitudinale analyses pas zinvol zijn na de verzameling van nieuwe data in metingen 4 en 5, en dat is inderdaad een goede reden dat er tot nu toe weinig echt longitudinale analyses zijn gedaan.

Omdat de onderzoekers bewoners uit de hele provincie Groningen hebben ondervraagd is het mogelijk de analyses uit te splitsen naar bewoners die wel en niet in aardbevingsgebied wonen. Hoewel er soms wel onderscheid in de analyses is gemaakt naar mate van schade, is er geen onderscheid gemaakt naar herkomstgebied van de respondent. Dit lijkt ons wel zinvol. De onderzoekers geven aan deze analyses overigens wel te hebben uitgevoerd, maar die niet te hebben gerapporteerd, omdat het onderscheid naar de mate van schade relevanter is (nl. er zijn grotere verschillen) voor de uitkomstmaten dan het herkomstgebied van de respondent.

Er worden in de nabije toekomst analyses gedaan met behulp van het Lifelines onderzoek. Het Lifelines onderzoek loopt sinds 2006 en is primair een onderzoek naar de gezondheid van Groningers, maar in dat onderzoek zijn ook vragen gesteld over bijvoorbeeld de psychische gesteldheid van personen. Het mooie van het Lifelines onderzoek is dat mensen al zijn gevolgd vóór 2012. Vanaf 2015 is via Lifelines bovendien een vragenlijst afgenomen die onder andere de beleving van aardbevingen en schade in kaart brengt. Zodoende kan er dus worden geanalyseerd hoe veranderingen in gezondheid in de laatste tien jaar samenhangen met de ondervonden schade, en kunnen bevindingen uit Lifelines en Gronings Perspectief worden gekoppeld. Deze manier van het onderzoeken van veranderingen binnen mensen, vóór en na de aardbevingen moet gezien worden als de Gouden Standaard als het gaat om niet-experimenteel effectonderzoek. De parallelle analyses van Lifelines en Groninger Perspectief worden in 2017 uitgevoerd en bieden een unieke mogelijkheid om 'effecten' van de gaswinning en aardbevingen te meten.

### **Bruikbaarheid voor toekomstige effectmetingen leefbaarheid**

Gronings Perspectief is bijzonder bruikbaar als blauwdruk voor het meten van de effecten van aardbevingen op leefbaarheid. De studie is van zeer hoge kwaliteit. De studie is longitudinaal opgezet, zodat door de tijd heen veranderingen in uitkomstmaten en mogelijke oorzaken en alternatieve verklaringen kunnen worden gemeten. Door de koppeling aan het Lifelines onderzoek is er daarnaast een unieke longitudinale vergelijkingsgroep (mits het Lifelines onderzoek wordt voortgezet – daarover bestaat twijfel), omdat het Lifelines onderzoek zowel binnen als buiten het aardbevingsgebied en de provincie Groningen een groot aantal mensen ondervraagt, en al gestart is voordat de huidige aardbevingsproblematiek begon. Er komen in het Gronings Perspectief veel verschillende onderwerpen aan bod. Daaronder valt leefbaarheid, hoewel het onderzoek zich met name richt op de psychosociale en gezondheidsgevolgen van gaswinning voor inwoners van het aardbevingsgebied. In samenspraak met de opdrachtgever is ervoor gekozen niet veel vragen over leefbaarheid op te nemen.

Voor het doen van toekomstige effectmetingen op leefbaarheid zou eerst moeten worden vastgesteld welk soort effecten men wil precies evalueren (bijvoorbeeld wat het doel of de uitkomstvariabelen zijn). Er bestaat een grote kans dat vragen gesteld in Gronings Perspectief informatie bieden over deze mogelijke effecten.

Een ander waardevol element van Gronings Perspectief is de combinatie van survey-onderzoek met kwalitatief onderzoek (een mixed-methods benadering). Dit maakt dat veranderingen in bijvoorbeeld de tevredenheid met de woonomgeving beter begrepen kunnen worden.

## **4.2 Wonen en aardbevingen in Groningen (OTB Delft)**

Het onderzoek Wonen en aardbevingen in Groningen is een groot onderzoek naar het wonen in Groningen door OTB Delft, in samenwerking met CMO STAMM. Het onderzoek omvat negen verschillende deelrapporten. De deelrapporten waarin specifieke analyses worden gedaan over de woningmarkt, of waar kwalitatief onderzoek is gebruikt zijn in onze review buiten beschouwing gelaten. Specifiek reviewen we in deze studie het deelrapport 'Wonen en aardbevingen in Groningen: een onderzoek in negen gemeenten', en het bijbehorende bijlagenboek. In dit deel van het onderzoek wordt respondenten vragen gesteld over de tevredenheid met de woning, woonomgeving, verhuisplannen, en hun reactie op aardbevingen. Het onderzoek is uitgevoerd in juni 2015.

### **Algemeen oordeel**

Dit is een goed opgezet onderzoek. Het steekproefdesign is goed gekozen. Het onderzoek laat ook goed zien dat er sprake is van een complexe dynamiek (bijvoorbeeld in verhuisbewegingen) binnen het aardbevingsgebied, die een breed opgezet onderzoek rechtvaardigt. Het onderzoek is niet longitudinaal uitgevoerd; we weten niet of de momentopname zoals die is genomen in 2015 nog steeds zou gelden. Wel geeft het onderzoek een goed beeld van de complexe dynamiek die er bestaat tussen krimp en het effect van aardbevingen op leefbaarheid en bijvoorbeeld verhuigeneidheid. Een belangrijk verbeterpunt is de hoogte van de respons. Van alle uitgenodigde mensen heeft 21% meegedaan. Hoewel er non-responsanalyses zijn uitgevoerd blijft er een risico bestaan op vertekening in de uitkomstvariabelen door deze non-respons. Een

ander verbeterpunt betreft de kwaliteit van de vragenlijst. Ook moet vermeld worden dat er een enorme berg aan data is verzameld in dit onderzoek, terwijl er niet veel analyses zijn uitgevoerd. De data zijn duidelijk onder-geanalyseerd, en verdiepende analyses naar de verschillen in ervaren leefbaarheid ontbreken.

Dit onderzoek is niet heel bruikbaar als blauwdruk voor toekomstig onderzoek naar effecten van beleid op leefbaarheid. In dit onderzoek is alleen gekeken naar bewoners van het aardbevingsgebied. Als het gaat om causale relaties is het moeilijk het beeld dat uit het onderzoek naar voren komt in een bredere regionale of landelijke context te plaatsen. Ook is het onderzoek niet longitudinaal van opzet. Hoewel het onderzoek wordt herhaald, gebeurt dit door het trekken van een verse steekproef, zodat veranderingen binnen individuen, die nodig zijn voor goede effectmetingen, niet mogelijk zijn.

### **Doelpopulatie**

De inwoners van negen gemeenten: Delfzijl, Eemsmond, Appingedam, Loppersum, Bedum, De Marne, Slochteren, Ten Boer en Winsum. Deze gemeenten samen vormen het 'aardbevingsgebied', hoewel dat in de toekomst door eventuele nieuwe bevingen kan verschuiven. Drie gemeenten (Delfzijl, De Marne en Eemsmond) liggen maar voor een klein deel binnen het huidige aardbevingsgebied. In het rapport wordt een onderscheid gemaakt tussen mensen binnen gemeenten die in- of buiten het aardbevingsgebied wonen.

### **Steekproeftrekking**

In totaal zijn ongeveer 19000 inwoners uitgenodigd om mee te doen.

Er zijn 811 inwoners uit het Groninger Panel aangeschreven (woonachtig in een van de negen gemeenten). Het panel (met destijds een omvang van 3000 leden) zou representatief zijn voor de bevolking van de provincie Groningen. Informatie over de samenstelling van het Groninger Panel is niet in het OTB-rapport beschikbaar (maar zie [sectie 4.4](#) voor een uitgebreidere beschrijving van het Groninger Panel).

De overige potentiële deelnemers zijn via een aselechte steekproef getrokken uit de Basisregistratie Personen (BRP) van de negen gemeenten. In vier gemeenten was sprake van een gestratificeerde steekproef, waarbij er oversampled werd in bepaalde kleine kernen, zodat er voldoende bewoners uit kleine kernen aan het onderzoek zouden deelnemen. Deze manier van steekproeftrekken lijkt op de steekproeftrekking zoals die in Groninger Perspectief ook is gehanteerd (zie [sectie 4.1](#)) en is een goede methode als een van de doelen van het onderzoek is om verschillen te onderzoeken tussen bewoners uit grotere en kleinere kernen.

### **Resultaat veldwerk**

De aselechte steekproef bestond uit 18436 personen, de deelnemers uit het Groninger Panel uit 811 personen.

Inwoners van de negen Groningse gemeenten ontvingen een uitnodigingsbrief met een URL naar de enquête, en daarnaast nog één herinnering. Bovendien konden inwoners op verzoek ook een schriftelijke versie van de enquête toegestuurd krijgen. Er waren geen beloningen voor respondenten. De respons bedroeg 20,8%; 3824 van de 18436 aangeschreven personen vulden de vragenlijst in. Dit is een relatief laag

responspercentage.

Bij de personen uit het Groninger Panel was de respons 65,2%; 529 van de 811 uitgenodigde panelleden vulden de enquête in. Bij de werving van het Groninger Panel is echter ook non-respons opgetreden. Deze non-respons is niet meegenomen in de analyses. Zou deze wel meegenomen worden, dan zou de respons uitkomen op ongeveer 13%.

### Correctie

Als gevolg van het gebruik van ongelijke selectiekansen en de non-respons van ongeveer 80% is er grote kans op vertekening in het onderzoek. Er is in dit onderzoek een complexe weegmethode gebruikt om te corrigeren voor ongelijke kansen in de steekproeftrekking en non-respons. In een eerste stap zijn er *designgewichten* berekend: weegfactoren die corrigeren voor de ongelijke kansen in de steekproeven, zodat in de resultaten de kansen voor elke persoon in de steekproef weer gelijk zijn. Het maken van designgewichten is een noodzakelijke stap voor een goede weegprocedure, die echter niet altijd wordt uitgevoerd in de andere onderzoeken die we in dit rapport reviewen. Om te corrigeren voor non-respons is er vervolgens gekeken naar vertekening op vier verschillende variabelen, en is er stap voor stap voor die variabelen gewogen. Allereerst is er gewogen naar huishoudtype (alleenstaand, samenwonend, met kinderen, etc.) en leeftijd. Daarna is er gewogen naar eigendomsverhouding (Syswov, het Systeem Woning Voorraad van de overheid), en vervolgens een tweede keer naar huishoudtype en eigendomsverhouding. Dit betreft een handmatige vorm van multiplicatief wegen waarin stapsgewijs voor meerdere variabelen tegelijkertijd wordt gewogen. In dit onderzoek is echter voor alle variabelen slechts één- of tweemaal gewogen, terwijl in een goede procedure zou moeten worden doorgegaan totdat de weegfactoren niet meer veranderen (zijn geconvergeerd). Multiplicatieve weging is in de basis wel een goede weegmethode en het is goed dat expliciet design- en non-responsgewichten apart zijn berekend. Als laatste stap is er gekeken of de mate van schade aan woningen in de studie overeenkwam met de schademeldingen in het schaderegister van de NAM. Dit bleek het geval.

De onderzoekers verwachtten dat door de gehanteerde weegprocedure de respons representatief zou worden met betrekking tot de andere relevante variabelen. De weegvariabelen die gebruikt worden lijken ons goed gekozen, vooral omdat dit onderzoek gaat om een woningmarktonderzoek, waarbij leefbaarheid als thema eigenlijk van secundair belang is. Variabelen als eigendomstype, type huishouden en leeftijd zijn sterk gerelateerd aan de belangrijke uitkomstvariabelen (wens tot en daadwerkelijk verhuisgedrag), en zijn daarmee effectieve weegvariabelen.

### Vragenlijst

De vragenlijst is opgenomen in het bijlagenboek van het rapport. De vragen die gesteld worden zijn voor een deel afkomstig uit de monitor Eemsdelta, die bij de start van dit onderzoek al liep. Andere vragen naar leefbaarheid zijn deels afkomstig uit CBS-onderzoek (met name het WoON-onderzoek, zie [sectie 4.3](#)), zodat vergelijkingen kunnen worden gemaakt tussen deze onderzoeken. Overige vragen zijn door de onderzoekers zelf opgesteld. Er wordt niet gerapporteerd of de vragenlijst getest is.

Het is goed om aan te sluiten bij bestaand onderzoek, zeker als dat onderzoek langdurig wordt gebruikt en vaker dient als vergelijkingsmateriaal. Daartegenover staat dat enquêtevragen altijd moeten passen bij de specifieke onderzoeksvraag en populatie. Tussen deze twee uitgangspunten bestaat altijd een zekere spanning. De vragenlijst van 'Wonen en aardbevingen in Groningen' is logisch opgebouwd, en kent een redelijke balans tussen feitelijke vragen en vragen over gevoelens en meningen. In sommige gevallen wordt er naar ons idee te veel naar de mening van de respondent gevraagd, zonder ook feitelijke vragen over diezelfde aspecten te stellen. Er wordt bijvoorbeeld specifiek gevraagd naar de tevredenheid met het voorzieningenniveau in de omgeving, en ontwikkelingen daarvan op een heel aantal aspecten (bijvoorbeeld: supermarkt, scholen, gezondheidszorg, sporten). Er wordt respondenten echter niet concreet gevraagd naar het gebruik van deze voorzieningen. De onderzoekers geven aan dat die keuze bewust gemaakt is, omdat de aanwezigheid van bijvoorbeeld een basisschool belangrijker is voor leefbaarheid, dan het daadwerkelijk gebruik daarvan.

Sommige vragen in het onderzoek zijn verder conditioneel, dat wil zeggen dat afhankelijk van eerdere antwoorden in de vragenlijst (bijvoorbeeld of mensen gaan verhuizen of niet) een specifieke vervolgvraag volgt (bijvoorbeeld waarom mensen wel of niet gaan verhuizen). Naar onze mening verschillen de conditionele vervolgvragen en antwoordcategorieën bij deze vragen onnodig van elkaar. Deze hadden meer vergelijkbaar kunnen zijn. Tenslotte vinden we de vragenlijst niet erg neutraal: In de introductie wordt bijvoorbeeld gesteld dat het onderzoek onder andere gaat over "of uw woning minder waard geworden [is] door de aardbevingen". Dit maakt de vragenlijst minder relevant voor bijvoorbeeld huurders, of mensen die zich niet druk maken om hun woningwaarde. In het algemeen vinden we dat de vragenlijst van voldoende kwaliteit is, maar dat er bij individuele vragen tekortkomingen zijn in de formuleringen van vragen en antwoordcategorieën.

### **Analyse**

Het is het vermelden waard dat in de appendix van het rapport ook een serie frequentietabellen is opgenomen met daarin een uitsplitsing van de resultaten per vraag naar sociaal-demografische achtergrond van de respondenten. Dit zorgt voor een transparant rapport en goede beschrijvende analyses.

De onderzoekers hebben ervoor gekozen om geen significantietoetsen uit te voeren, want 'die worden alleen bij puur wetenschappelijk onderzoek gebruikt'. Op zich begrijpen we de keuze van de onderzoekers om een niet-technisch rapport op te willen leveren, maar als er vergelijkingen worden gemaakt (tussen groepen of door de tijd), dan is het naar ons idee wel noodzakelijk om toetsen en onzekerheidsmarges te rapporteren. Zo krijgt de lezer een betere indruk of een gerapporteerde verandering of een gerapporteerd verschil ook echt als zodanig kan worden geïnterpreteerd of wellicht alleen het gevolg is van toevallige fouten of ruis als gevolg van het werken met een steekproef.

De onderzoekers hebben op meerdere manieren geprobeerd de bevindingen te duiden. Zo hebben ze vrij uitgebreid kwalitatief onderzoek gedaan met behulp van bewonersavonden en grote focusgroepen, om bevindingen uit het vragenlijstonderzoek uit te diepen en beleidsimplicaties te bepalen. Ook zijn er analyses uitgevoerd op het Sociaal Statistisch Bestand, om de daadwerkelijke verhuisbewegingen van de

Groningers in kaart te brengen en te koppelen aan de wens tot verhuizen die aan bod komt in de survey. Hoewel we in deze review alleen het deelrapport beoordelen dat aan de hand van surveydata tot stand is gekomen, zijn er in dit onderzoek heel veel data verzameld, bijvoorbeeld in focusgroepen, en over de verhuisbewegingen van en naar Groningen. De data zijn erg rijk en bieden tal van mogelijkheden voor verdiepende analyses, die tot nu toe niet zijn uitgevoerd. Met andere woorden: De data in dit onderzoek zijn ondergeanalyseerd.

### **Bruikbaarheid voor toekomstige effectmetingen leefbaarheid**

Dit onderzoek is niet longitudinaal van opzet. Hoewel er plannen zijn dit onderzoek in 2017 te herhalen, zullen daar niet dezelfde mensen opnieuw worden ondervraagd. Dit maakt het lastig om veranderingen in de tijd in uitkomstmaten en mogelijke oorzaken en alternatieve verklaringen goed te meten. Een onderdeel uit dit onderzoek dat wel heel bruikbaar is, is de steekproefopzet. In de steekproef krijgen personen dichterbij de kern van het aardbevingsgebied een hogere selectiekans, waardoor in die gebieden met meer precisie uitspraken kunnen worden gedaan. Ook bruikbaar is de combinatie van survey-onderzoek met kwalitatief onderzoek (mixed-methods benadering) en de koppeling met het Sociaal Statistisch Bestand.

## **4.3 Leefbaarometer 2.0/WoON**

De Leefbaarometer is een instrument om de leefbaarheid in Nederland te schatten op een zeer laag schaalniveau (zes-cijferige postcodes – eigenlijk vier cijfers en twee letters). De Leefbaarometer bestaat sinds 2008. In 2015 is de Leefbaarometer vernieuwd, vandaar de toevoeging 2.0 in de naam. De leefbaarheid wordt geschat op basis van allerlei omgevingscondities. Dit instrument betreft dus heel Nederland en niet specifiek Groningen. De Leefbaarometer bestaat uit een model dat allerlei indicatoren gebruikt om per zes-cijferig postcodegebied een schatting van de leefbaarheid te geven:

- Er is een model gebouwd dat het oordeel van de mensen over de leefbaarheid in hun omgeving probeert te verklaren uit een grote set indicatoren. Het gaat om ongeveer honderd fysieke kenmerken van woningen, aangevuld met subjectieve gegevens afkomstig uit het WoonOnderzoek Nederland (WoON) dat periodiek wordt uitgevoerd door het Centraal Bureau voor de Statistiek.
- Er is een model dat de huizenprijzen probeert te verklaren uit een grote set indicatoren: De hoogte van de huizenprijzen wordt dus gezien als een indirecte indicator voor de leefbaarheid.

Beide modellen worden gecombineerd tot één model voor de leefbaarheid. De verklarende variabelen in beide modellen zijn afkomstig uit integrale registraties, zoals Basisregistratie Personen (BRP), de Basisadministratie Gebouwen (BAG) en politiegegevens. Een van de variabelen is een indicator voor het aardbevingsrisico. Die komt uit de Risicokaart. Er kan dus worden nagegaan in hoeverre het aardbevingsrisico van invloed is op de leefbaarheid.

De twee modellen worden geschat voor de gebieden waarvoor gegevens uit WoON beschikbaar zijn. Vervolgens kunnen die modellen worden gebruikt om de leefbaarheid te schatten in gebieden waar geen WoON-gegevens beschikbaar zijn. Omdat de Leefbaarometer uit twee afzonderlijke delen bestaat die relevant zijn voor deze review, hebben we ervoor gekozen de verdere bespreking uit te splitsen. Eerst zullen we het

WoON onderzoek apart behandelen, en daarna bespreken hoe gegevens van WoON, samen met andere data worden gebruikt in de Leefbaarometer.

### 4.3.1 WoonOnderzoek Nederland (WoON)

Het WoonOnderzoek Nederland (WoON) verzamelt statistische informatie over de woonsituatie van de Nederlandse bevolking en haar wensen en behoeften op het gebied van wonen. Het WoON wordt driejaarlijks uitgevoerd door het CBS, waardoor over langere periodes kan worden gekeken hoe de woonsituatie en wensen veranderen door de tijd heen. Daarbij wordt onder andere aandacht besteed aan de samenstelling van huishoudens, de woning en woonomgeving, de woonlasten, woonwensen en verhuizingen. Het subjectieve deel van leefbaarheid wordt hierbinnen gemeten door middel van drie vragen.

#### *Algemeen oordeel*

Het WoON is een landelijk dekkend, grootschalig en periodiek onderzoek naar de woonomgeving van mensen in Nederland. In het WoON worden maar enkele vragen gesteld die te maken hebben met het thema leefbaarheid, wat een duidelijke beperking is van dit onderzoek ten opzichte van de twee hiervoor besproken onderzoeken. Een sterk punt van het WoON onderzoek is de hoge kwaliteit van het veldwerk, dat tot uiting komt in een hoge respons en goede correctiemethode voor non-respons. Een ander belangrijk sterk punt is het feit dat het onderzoek het mogelijk maakt om regio's binnen Nederland te vergelijken. Deze vergelijkingen zijn mogelijk op provinciaal en regionaal niveau. Voor een vergelijking op gemeentelijk niveau binnen de provincie Groningen is de steekproefgrootte te klein. Het WoON onderzoek is bruikbaar voor toekomstige effectmetingen, omdat door middel van langdurige tijdreeksen een goed algemeen beeld kan worden gegeven van leefbaarheid op regionaal niveau. Er worden te weinig vragen over leefbaarheid gesteld om een veelzijdig beeld van leefbaarheid te krijgen, en daarom is dit onderzoek in de praktijk vaak niet voldoende geschikt om effecten op meerdere dimensies van leefbaarheid te kunnen meten.

#### *Onderzoekspopulatie*

De onderzoekspopulatie bestaat uit personen van 18 jaar en ouder in particuliere huishoudens in Nederland. De doelpopulatie is dus niet beperkt tot Groningen of een deel van Groningen. In de WoON-rapportage wordt niet ingegaan op de situatie in Groningen, maar dit onderzoek zou daarvoor wel gebruikt kunnen worden.

#### *Steekproeftrekking*

De steekproef voor het WoON wordt getrokken uit de Nederlandse bevolking van 18 jaar of ouder die geregistreerd is in de Basisregistratie Personen (BRP). De steekproef is een gestratificeerde steekproef met landelijke dekking van gemeenten. Enkele gemeenten hebben hun steekproef laten vergroten ('over-sampling'), zodat zij ook gegarandeerd zijn van betrouwbare uitspraken op gemeentenniveau. De via enquêtering verzamelde gegevens worden aangevuld met gegevens uit registers.

#### *Veldwerk*

Het WoON-onderzoek wordt afgenomen aan de hand van persoonlijke interviews, telefonische interviews en sinds 2009 ook via internet. Het aantal respondenten moet minimaal 60.000 bedragen. De respons op WoON 2015 was 57%.

### *Correctie*

In WoON wordt een uitgebreide weging toegepast op basis van geslacht, leeftijd, herkomst, inkomen, WOZ, positie in huishouden, omvang huishouden, eigendom woning, en woontype. Het CBS, dat het veldwerk uitvoert voor het WoON, heeft toegang tot registervariabelen die het mogelijk maken een specifiek weegmodel te bouwen dat zowel de variatie in responskansen verklaart, en samenhangt met de verschillende doelvariabelen (zoals uitgelegd in [Hoofdstuk 3](#) en [Bijlage 4](#)). Daardoor is het mogelijk om op een adequate manier te corrigeren voor vertekening door non-respons die tijdens het veldwerk is ontstaan. Het is van belang te melden dat het CBS in deze zin een bevoorrechte positie heeft ten opzichte van andere onderzoekers, omdat voor het CBS veel meer populatiegegevens beschikbaar zijn.

### *Vragenlijst*

Voor het WoON onderzoek wordt een zogenaamde mixed-mode dataverzameling gebruikt. Respondenten worden, afhankelijk van hun achtergrond, benaderd via internet. Non-respondenten worden daarna telefonisch benaderd, als er een telefoonnummer van het huishouden bekend is, of anders met behulp van interviewers thuis geïnterviewd. De uitnodigingsbrieven voor het WoON zijn neutraal geformuleerd, en datzelfde geldt voor de vragenlijst. In de vragenlijst van WoON worden allerlei vragen gesteld over de leefomgeving, de kwaliteit van de woning, voorzieningen in de buurt en eventuele verhuishwensen. Er worden drie specifieke vragen gesteld over leefbaarheid, namelijk:

- Hoe tevreden bent u met uw woonomgeving?
- Stelling: Het is vervelend in deze buurt te wonen.
- Stelling: Ik voel me thuis in deze buurt.

### *Analyse*

De drie bovenstaande vragen worden in veel rapportages samengevoegd tot één indicator die 'tevredenheid met de woonomgeving' meet. Omdat het WoON landelijk wordt afgenomen, en een grote steekproef heeft, kunnen voor relatief kleine geografische gebieden (regio's) vergelijkingen worden gemaakt in de mate van tevredenheid met de woonomgeving. Omdat het WoON onderzoek ook periodiek wordt herhaald kunnen zowel tussen regio's als door de tijd vergelijkingen worden gemaakt. Voor de provincie Groningen en het aardbevingsgebied in het bijzonder is van belang dat de grootte van de steekproef niet zodanig is dat er op het niveau van dorp, of zelfs gemeente goede analyses kunnen gedaan. Voor het vergelijken van COROP-regio's (Delfzijl en omgeving, Oost-Groningen, Overig Groningen) biedt WoON wel goede data. Het CBS voert zelf alleen beschrijvende analyses, en geen multivariate analyses uit om bijvoorbeeld verschillen in subjectieve leefbaarheid te verklaren.

### *Bruikbaarheid voor toekomstige effectmetingen leefbaarheid*

Het WoON onderzoek is bruikbaar voor toekomstige effectmetingen, omdat door middel van langdurige tijdreeksen een goed algemeen beeld kan worden gegeven van leefbaarheid op regionaal niveau. Wel wordt er telkens een nieuwe steekproef getrokken, zodat veranderingen binnen personen niet kunnen worden gemeten. Er worden te weinig vragen over leefbaarheid gesteld om een veelzijdig beeld van leefbaarheid te krijgen, en daarom is dit onderzoek in de praktijk vaak niet voldoende geschikt om effecten op meerdere dimensies van leefbaarheid te meten. Het is van belang te vermelden dat gemeenten in Nederland de mogelijkheid hebben om in hun gemeente een extra steekproef te trekken, zodat resultaten uit het WoON voor die

gemeente beschikbaar zijn.

### 4.3.2 Leefbaarometer 2.0

De Leefbaarometer 2.0 bestaat uit de hierboven besproken subjectieve indicatoren voor leefbaarheid afkomstig uit het WoON, gecombineerd met allerlei gegevens uit registers die per postcodegebied beschikbaar zijn. Waar het WoON qua omvang van de steekproef groot genoeg is om op regionaal niveau de leefbaarheid in kaart te brengen, ambieert de Leefbaarometer dit te doen op straat/buurniveau (zes-cijferig postcodegebied).

Aan de hand van de fysieke indicatoren wordt een model gebouwd, dat voorspellingen doet wat betreft de leefbaarheid per postcodegebied. Het gaat om ongeveer honderd fysieke kenmerken van postcodegebieden, die uiteenvallen in kenmerken van de woningen (26), bewoners (16), voorzieningen (23), veiligheid van de woonomgeving (6) en de fysieke omgeving (29).

Naast een model voor de leefbaarheid is er ook een model dat de huizenprijzen probeert te verklaren uit een grote set indicatoren: De hoogte van de huizenprijzen worden daarbij gezien als een indirecte indicator voor de leefbaarheid. Bij het lezen van het nu volgende oordeel over de Leefbaarometer is het van belang te noemen dat de fysieke indicatoren niet gecreëerd zijn op basis van vragenlijstdata. De onderdelen 'doelpopulatie', 'steekproef' en 'correctie' uit onze meetlat zijn derhalve minder van belang omdat de Leefbaarometer data beschrijft van alle postcodes in Nederland. Deze onderdelen beschrijven we daarom kort. Wel is van belang hoe er aan de hand van de fysieke indicatoren Leefbaarheidsscores worden berekend.

#### *Algemeen oordeel*

Sterk aan de Leefbaarometer is dat deze op een kleinschalig niveau (zes-cijferig postcodegebied) een leefbaarheidsindex bepaalt. Het onderzoek wordt periodiek herhaald, zodat de ontwikkelingen in de leefbaarheidsindex in het aardbevingsgebied kunnen worden vergeleken met ontwikkelingen in andere gebieden. Om tot preciezere schattingen te komen, zouden gemeenten in het aardbevingsgebied kunnen vragen om in toekomstig WoON in hun gemeente te oversamplen. Een beperkende factor is dat de subjectieve scores gebaseerd zijn op slechts drie vragen die vragen naar de algemene tevredenheid met de woonomgeving. In het WoON-onderzoek zijn niet meer directe vragen beschikbaar die expliciet over de leefbaarheid gaan, maar er zijn binnen andere grote landelijke onderzoeken zoals de Veiligheidsmonitor, meer subjectieve indicatoren van leefbaarheid beschikbaar om tot een wat gebalanceerder meetinstrument te komen. Het meetinstrument voor leefbaarheid dat nu wordt gebruikt is niet goed van toepassing op Groningen. De voorspelde scores uit de Leefbaarometer zijn namelijk hoger dan de scores op basis van alleen de subjectieve vragen uit WoON.

Voor toekomstig onderzoek naar de effecten van beleid op leefbaarheid is dit onderzoek op ongeveer dezelfde wijze bruikbaar als het WoON. Een extra voordeel van de Leefbaarometer is dat de leefbaarheid op buurniveau kan worden vergeleken door de tijd heen. Een nadeel ten opzichte van WoON is dat veel fysieke indicatoren maar in beperkte mate leefbaarheid meten.

### *Doelpopulatie*

Omdat de Leefbaarometer zich baseert op alle postcodegebieden in Nederland is er geen sprake van een echte doelpopulatie. Naast de al eerder besproken subjectieve indicatoren van leefbaarheid uit WoON worden ongeveer honderd fysieke indicatoren gebruikt op 6-cijferig postcodegebied. Deze indicatoren zijn uit te splitsen naar de volgende domeinen: type woningen, fysieke omgeving van woning, nabijheid voorzieningen, bevolkingssamenstelling omgeving, criminaliteitscijfers, huizenprijzen. De verhouding tussen objectieve en subjectieve indicatoren helt in alle modellen erg door naar objectieve indicatoren.

### *Steekproef*

Leefbaarometer 2.0 is gebaseerd op bijna 83000 waarnemingen uit het WoON, gecombineerd met fysieke gegevens die beschikbaar zijn voor ieder postcodegebied in Nederland (zes-cijferige postcode).

### *Correctie*

De gecorrigeerde resultaten uit het WoON worden gebruikt. Voor de fysieke gegevens is er geen sprake van non-respons of correcties.

### *Vragenlijst*

Voor de subjectieve indicatoren worden de drie vragen uit het WoON gebruikt (zie vorige sectie). Voor de fysieke indicatoren wordt gebruikt gemaakt van zoveel mogelijk fysieke gegevens die beschikbaar zijn in registers.

### *Analyse*

De leefbaarheid per zes-cijferig postcodegebied is een modelmatige schatting. In deze aanpak zitten minimaal twee soorten onzekerheden: 1) steekproeffouten van WoON en 2) het risico van foute modelveronderstellingen. Het totale model om leefbaarheid te meten bestaat uit ongeveer honderd indicatoren, waaronder drie afkomstig uit vragenlijstonderzoek (zie hierboven). De onderzoekers van de Leefbaarometer zijn zich bewust van de enigszins willekeurige selectie van fysieke en subjectieve indicatoren van leefbaarheid. In de analyses van de data wordt daarom een onderscheid gemaakt tussen twee modellen: een model met alleen fysieke kenmerken, en een model met zowel fysieke kenmerken als de drie subjectieve vragen uit WoON.

Het model met de honderd fysieke factoren correleert ongeveer 0,85 met de stellingen over leefbaarheid uit het WoON. Deze correlatie is hoog, en geeft aan dat de fysieke factoren sterk positief samenhangen met de subjectieve tevredenheid van inwoners met hun woonomgeving. Belangrijk is wel te melden dat er een aantal uitbijters is, waarvan de fysieke en subjectieve scores niet overeenkomen. Twee daarvan zijn de gebieden Delfzijl e.o. en Oost-Groningen. Volgens de auteurs van de Leefbaarometer kunnen de afwijkende scores in deze gebieden ontstaan door vergrijzing (Delfzijl) en aardbevingen (Noord Oost-Groningen). Hoewel voor heel Nederland de subjectieve leefbaarheid goed te voorspellen is aan de hand van de fysieke indicatoren, geldt dit dus niet voor delen van het aardbevingsgebied. Ook lijkt het zo te zijn dat de onderzoekers op een vrij exploratieve wijze de leefbaarheidsscores hebben gemaakt. Als dat zo is, kan de correlatie tussen fysieke factoren en subjectieve tevredenheid in een volgend Leefbaarometer-onderzoek lager uitvallen.

Er heeft externe validatie plaatsgevonden door de uitkomsten van de Leefbaarometer 2.0 voor te leggen aan lokale deskundigen. Daarbij waren ook deskundigen uit de Groningse gemeente Eemsum. Het oordeel van deze analyse was dat de schattingen van leefbaarheid niet altijd overeenkomen met de oordelen van deskundigen. Meestal wel, maar op wijkniveau wordt de leefbaarheid soms over- of onderschat.

#### *Bruikbaarheid voor toekomstige effectmetingen leefbaarheid*

De Leefbaarometer 2.0 is net als het WoON onderzoek bruikbaar voor effectmetingen, omdat het periodiek wordt uitgevoerd en landelijk dekkend is. Omdat er in de Leefbaarometer 2.0 meer informatie beschikbaar is per postcode dan in het WoON onderzoek, kunnen op kleinere schaal gebieden met elkaar worden vergeleken. Delen van het aardbevingsgebied in het bijzonder kunnen worden vergeleken met delen die buiten het aardbevingsgebied liggen. En het hele aardbevingsgebied kan worden vergeleken met andere regio's die in bepaalde opzichten op deze regio lijken. Het is wel van belang dat er daarbij een goed model wordt gespecificeerd dat recht doet aan de context van zowel Groningen als de vergelijkingsgebieden.

## 4.4 Leefbaarheid en bevolkingskrimp in Groningen (Sociaal Planbureau Groningen)

Het onderzoek 'Leefbaarheid en bevolkingskrimp in Groningen' is een onderdeel van de leefbaarheidsmonitor Groningen, die wordt uitgevoerd door het Sociaal Planbureau Groningen. Deze monitor volgt de ontwikkelingen en trends in de leefbaarheid in Groningen. Naast de trends en ontwikkelingen van de objectieve leefbaarheid volgt de leefbaarheidsmonitor ook de ervaren leefbaarheid. De ervaren leefbaarheid wordt hierbij in kaart gebracht via het Groninger Panel. Het Groninger panel is een online panel dat respondenten circa vijf keer per jaar vragen voorlegt over uiteenlopende onderwerpen die in Groningen spelen (bijvoorbeeld: verkeer, veiligheid, participatie, voorzieningen, of de zorg). In 2016 is het Groninger Panel bevraagd over de ervaren leefbaarheid en de waardering van de eigen woonomgeving. Deze vragenlijst is, naast in het Groninger Panel, ook uitgezet onder panelleden van het Fries Panel en het Drents Panel. Het is de bedoeling dat dit onderzoek elke twee jaar herhaald wordt. Specifiek reviewen we het Groninger Panel onderzoek uit 2016 over de ervaren leefbaarheid en waardering van de woonomgeving.

### **Algemeen oordeel**

Potentieel sterk aan dit onderzoek is dat het samen met de Leefbaarometer een van de weinige onderzoeken is die al zijn gestart voor of rond het begin van de huidige aardbevingsproblematiek in 2012. Daarnaast is een sterk punt dat er vergelijkingen te maken zijn met de provincies Friesland en Drenthe. Hoewel die provincies hun eigen dynamiek hebben als het gaat om bijvoorbeeld krimp, kan systematisch en langdurig worden bekeken in hoeverre leefbaarheidspercepties in Groningen, en daarbinnen specifieke gebieden, afwijken, omdat vragenlijsten over leefbaarheid ook in die provincies worden afgenomen. Daarnaast zou met behulp van het Groninger Panel de situatie voor, tijdens en straks (hopelijk) na de aardbevingen mogelijk kunnen worden geanalyseerd. De vragenlijst van dit onderzoek is uitgebreid. Van alle vijf de onderzoeken biedt deze vragenlijst het breedste beeld van leefbaarheid en onderliggende indicatoren.

Verbeterpunten zijn de kwaliteit van de rapportages, de hoge non-respons, de analyse en correctie voor non-respons, en de frequentie waarmee onderzoek wordt gedaan. Er is in het Groninger Panel voor gekozen om niet te frequent vragenlijstonderzoek te doen (om de maand een korte lijst). Er zou ruimte zijn om het panel intensiever te benutten, zodat er meer longitudinale data beschikbaar is om ontwikkelingen in Groningen te kunnen monitoren.

Deze studie is bruikbaar voor toekomstig onderzoek naar leefbaarheid. Het onderzoek is longitudinaal, zodat kan worden geanalyseerd voor welke soort mensen de gerapporteerde leefbaarheid verbeterd of verslechterd. Omdat de vragenlijst veel aspecten van leefbaarheid meet, zou(den delen van) deze lijst goed kunnen worden gebruikt om leefbaarheid te meten.

### **Doelpopulatie**

De doelpopulatie beslaat de Groninger bevolking van 18 jaar en ouder. Er wordt in dit onderzoek en de rapportage een onderscheid gemaakt naar krimpgebieden en niet-krimpgebieden, en het aardbevingsgebied en niet-aardbevingsgebied. De publieke documentatie van de onderzoeken van het Groninger Panel is niet zo uitgebreid als die van de andere onderzoeken, en precieze informatie over de steekproeftrekking ontbreekt. Deze informatie hebben we mondeling verkregen in een gesprek met de onderzoekers.

### **Steekproeftrekking**

Het Groninger Panel streeft ernaar een representatieve afspiegeling te zijn van de Groninger bevolking van 18 jaar en ouder uit alle 23 Groninger gemeenten. Er wordt naar gestreefd om minimaal 1% van de bevolking uit iedere gemeente in het panel te krijgen. Dit criterium wordt in 18 Groningse gemeenten ongeveer gehaald, in de overige gemeenten net niet.

Het panel is eind 2012 opgericht, en in maart 2013 zijn de eerste vragenlijsten afgenomen. Mensen kunnen zich niet zelf aanmelden voor het panel; het betreft een panel zonder zelfselectie. De panelwerving in 2012 is verlopen via Cendris. Cendris beheert een database met alle postadressen in Nederland. 10.000 huishoudens zijn aangeschreven, en 1700 mensen zijn lid geworden van het panel. Na 2012 is er ieder jaar bijgeworven. Respondenten worden in de bijwerving telkens met enkelvoudig aselechte steekproeven geworven uit de Basisregistratie Personen (BRP). De grootte van bijwerving fluctueert, en de respons ook (maar was ongeveer 20%). Tenslotte zijn soms mensen geworven uit andere (offline) onderzoeken van het Sociaal Planbureau Groningen. Op dit moment zijn er 4700 panelleden. Er worden geen specifieke pogingen gedaan om bijvoorbeeld verhuizers naar en uit Groningen te identificeren. Al het onderzoek wordt online gedaan. Bij de vragenlijst naar Leefbaarheid en Bevolkingskrimp in Groningen, maar ook andere vragenlijsten, krijgen respondenten een e-mail met een uitnodiging naar de vragenlijst en een herinnering. Respondenten krijgen geen geldelijke beloning voor het invullen van de vragenlijst. Periodiek wordt er een verloting gedaan van een aantal cadeaubonnen.

Er wordt gestreefd naar een responspercentage van 50% van mensen uit het panel dat ruimschoots wordt gehaald. Het is wel belangrijk te bedenken dat deze respons 50% betreft van de leden van het panel. Conditioneel op de bruto steekproef bedraagt

de respons dus ongeveer 10% (20% in de werving, en 50% bij iedere vragenlijst). Er zijn analyses gedaan op non-respons in het panel. Deze non-respons bestaat dus uit non-respons in de werving, en bij iedere vragenlijst. Leeftijd is een voorspeller van non-respons. Mensen in de leeftijd van 50-75 zijn oververtegenwoordigd, terwijl jongeren (18-25) ondervertegenwoordigd zijn. Allochtonen, lager opgeleiden en plattelandsbewoners zijn ook ondervertegenwoordigd.

### **Resultaat veldwerk**

Van het totale aantal panelleden in 2016 (ongeveer 3800) hebben bijna 2000 respondenten meegedaan aan het onderzoek naar leefbaarheid en krimp. Dit betekent dat ongeveer 50% van de panelleden heeft meegedaan. Het is belangrijk om te bedenken dat er in het Groninger Panel sprake is van een hoge non-respons. Door de hoge non-respons in de werving van het panel (80%) en non-respons bij iedere vragenlijst (50%) ligt de totale non-respons op ongeveer 90% (11% respons). Er bestaat daarom een risico dat non-respondenten verschillen van respondenten op allerlei kenmerken die niet zijn gemeten.

### **Correctie**

Het Groninger panel is niet representatief voor 'Groningen'. Er is namelijk een steekproef getrokken per gemeente, en er zijn correcties nodig om iets te kunnen zeggen over de provincie of deelgebieden. Hoewel dat niet gerapporteerd wordt, zijn de data volgens de onderzoekers wel gecorrigeerd voor deze ongelijke selectiekansen.

De representativiteit van dit onderzoek kan op twee manieren zijn aangetast: (1) Bij de werving van het panel zal een behoorlijke hoeveelheid non-respons optreden, en (2) bij de peiling uit het panel is er ook een flinke hoeveelheid non-respons (56%). Er is door de onderzoekers gekeken naar algemene non-respons in het panel, maar er wordt in analyses alleen gewogen op de variabelen leeftijd en urbanisatiegraad (stad vs. platteland). Onduidelijk is of dit de juiste weegvariabelen zijn: Deze variabelen hangen waarschijnlijk niet sterk samen met ervaren leefbaarheid. Over de tijd heen zijn er ook mensen uitgevallen uit het panel, maar daar zijn geen specifieke analyses of correcties voor uitgevoerd. Er wordt telkens aselekt bijgeworven om de grootte van het panel op peil te houden en verder uit te bouwen. De correctiemethodes voor dit onderzoek zijn basaal. Een meer uitgebreide non-responsanalyse, en bijbehorende non-responscorrectie is nodig om met enige zekerheid iets te kunnen zeggen over de zekerheid van de uitkomsten.

### **Vragenlijst**

De vragenlijst is niet opgenomen in de onderzoeksverantwoording. Na een verzoek is de vragenlijst, de uitnodigingsbrief en andere informatie over de uitvoering van het onderzoek wel beschikbaar gekomen voor onze review. De uitnodigingsbrief is begrijpelijk en neutraal geformuleerd. De vragenlijst van het onderzoek vertoont gelijkenissen met het onderzoek 'Wonen en aardbevingen in Groningen', zoals besproken in [sectie 4.2](#). De reden hiervoor is dat dezelfde organisatie (CMO STAMM, waar het Sociaal Planbureau Groningen deel van uitmaakt) betrokken is geweest bij de uitvoering van beide onderzoeken. De vragenlijst richt zich op het meten van het construct 'leefbaarheid'. Van alle vijf onderzoeken die in dit rapport worden gereviewd, is dit het enige onderzoek dat expliciet en als voornaamste doel heeft leefbaarheid zo goed mogelijk te meten. Om deze reden is de vragenlijst uitgebreid, en wordt naar allerlei aspecten van leefbaarheid

gevraagd: tevredenheid met de woonomgeving, eigen inzet voor het behouden/versteken van de leefbaarheid, tevredenheid met het voorzieningenniveau, en een mogelijke reactie op het verdwijnen van deze voorzieningen. Het Sociaal Planbureau Groningen heeft op zijn website een schematisch model staan waarin de brede definitie van leefbaarheid en bijbehorende indicatoren (hoe is het goed wonen, leven, werken) uiteen wordt gezet.

De vragenlijst bestaat veelal uit stellingen waar de respondenten op reageren (bijvoorbeeld: “In hoeverre bent of het eens met (...)”, “In hoeverre vindt u het belangrijk dat (...)”). Zo wordt de respondent bijvoorbeeld gevraagd in hoeverre de volgende voorzieningen “in uw dorp of wijk belangrijk zijn”, met als zeven subvragen: [supermarkt, basisschool, huisarts, dorps huis, openbaar vervoer, pinautomaat]. Na deze vragen krijgen respondenten vervolgvragen over welke voorzieningen er zijn of zijn geweest, welke voorzieningen er missen, in welke mate respondenten het jammer zouden vinden als voorzieningen verdwijnen, en wat de respondent zou doen als een specifieke voorziening dreigt te verdwijnen. Hoewel het begrip leefbaarheid veel breder is dan tevredenheid met de woonomgeving en het voorzieningenniveau, bevat deze vragenlijst een groot aantal vragen rond het begrip leefbaarheid. In onze conclusie en aanbevelingen komen we terug op hoe het construct ‘leefbaarheid’ gemeten zou moeten en kunnen worden. De vragenlijst van het Groninger panel is in ieder geval waar het gaat om het meten van leefbaarheid het breedst opgezet van de door ons beoordeelde onderzoeken. De vele stellingen nodigen wel uit tot strategisch antwoordgedrag van de respondent; de respondent kan ervoor kiezen om de vragenlijst snel af te maken door telkens hetzelfde antwoord te kiezen (ook wel bekend als straightlining, zie [sectie 3.3](#)). Bij sommige vragen wordt “weet niet” als antwoordcategorie aangeboden, en bij andere vragen niet. Hoewel daar goede redenen voor kunnen zijn, blijft onduidelijk waarom deze keuzes per vraag zijn gemaakt.

## Analyse

In analyses wordt een onderscheid gemaakt naar krimp- en niet-krimp gemeenten. Er worden nergens onzekerheidsmarges vermeld. Het is daardoor lastig om verschillende uitkomsten met elkaar te vergelijken. Zo is de waardering voor de leefbaarheid in de krimpgebieden gelijk aan 7,4, terwijl de waardering buiten de krimpgebieden gelijk is aan 7,7. Is dit een significant verschil of is het slechts de ruis van de steekproef? Het is daarnaast hoogst onzeker of het verschil in leefbaarheid hetzelfde zou zijn na een goede correctieprocedure voor non-respons. Er bestaat een gerede kans dat het verschil tussen krimpgebieden en niet-krimpgebieden groter of kleiner zou zijn.

In de analyses met behulp van het panel worden weinig longitudinale analyses gedaan, hoewel dat wel zou kunnen met de data die er liggen. Er wordt elke twee maanden een kort onderzoek gedaan, en af en toe een langduriger onderzoek. Vragenlijstmodules worden elke twee jaar herhaald, als het gaat om het onderwerp leefbaarheid. Er zijn dus nog weinig echt longitudinale data, waarbij naar veranderingen wordt gekeken binnen dezelfde personen. Wel wordt er naar trends gekeken op geaggregeerd niveau, en worden er vergelijkingen gemaakt met Friesland en Drenthe, waar sinds kort ook planbureaus zijn, en waar min of meer dezelfde vragenlijsten worden gebruikt.

### **Bruikbaarheid voor effectmetingen leefbaarheid**

Het Groninger panel is zeer bruikbaar voor toekomstige metingen. Het Sociaal Planbureau Groningen doet longitudinaal onderzoek, hoewel er tot nu toe nog niet vaak vragenlijsten zijn herhaald. Over het onderwerp leefbaarheid is een goede, breed opgezette vragenlijst beschikbaar. Er is een vergelijkingsgroep doordat veel vragenlijsten gelijktijdig worden afgenomen in soortgelijke panels in Drenthe en Friesland. De grootte van het panel, en de hoeveelheid personen daarin afkomstig uit het aardbevingsgebied leent zich ervoor om het aardbevingsgebied te vergelijken met het niet-aardbevingsgebied binnen de provincie Groningen. Binnen het aardbevingsgebied zijn er te weinig deelnemers om op kleinere geografische schaal analyses te doen.

## **4.5 Onderzoek naar de tevredenheid met de schadeafhandeling door de NAM (APE Public Economics)**

De NAM heeft lange tijd een belangrijke rol gespeeld in het afhandelen van schadegevallen die door inwoners van het aardbevingsgebied gemeld zijn. Een goede afhandeling van deze schademeldingen is van groot belang voor het vertrouwen van burgers in de NAM en in de overheid. De Dialoogtafel Groningen heeft onderzoeksbureau APE in 2014/2015 de opdracht gegeven de tevredenheid van bewoners met de schadeafhandeling door de NAM te meten. Daartoe is een onderzoek opgezet waar behalve vragenlijsten ook kwalitatieve interviews en deskresearch zijn gebruikt. In deze review focussen we op het vragenlijst-onderzoek. Hierin stond de tevredenheid over 1) het meldproces van schade 2) de taxatie door een expert 3) de overeenstemming over de ernst, type en bedrag van de schade, 4) de afhandeling van de schade, en 5) in voorkomende gevallen tevredenheid met de klachtenprocedure van de NAM, centraal.

### **Algemeen oordeel**

Er is in dit onderzoek in theorie gekozen voor een slimme onderzoeksofzet. Door de steekproef te trekken onder alle mensen die een schade hebben gemeld bij de NAM kan direct worden geëvalueerd in welke mate gedupeerden van de aardbevingen in Groningen tevreden of ontevreden zijn met de manier waarop hun schademelding is afgehandeld. Er wordt niet uitgebreid naar leefbaarheid gevraagd, maar dat was ook niet de doelstelling van het onderzoek.

De grootste kanttekening bij dit onderzoek in het kader van onze review naar leefbaarheidsstudies is dat het alleen iets zegt over de afgehandelde schades in een relatief korte periode. Behalve dat er binnen het onderzoek wat onduidelijkheden zijn rond non-respons en correctie daarvoor, kunnen de volgende oorzaken leiden tot een vertekend beeld over de tevredenheid over de afhandeling van schades in het algemeen:

- Schades worden niet gemeld bij de NAM en staan dus niet in het register, dan wel omdat dit gaat om huurwoningen, omdat bewoners geen vertrouwen hebben in de NAM, of omdat de schade ergens anders is gemeld.
- Schades worden niet juist gemeld/niet in behandeling genomen door de NAM.
- Schademeldingen lopen nog. Dat wil zeggen dat sommige mensen hun schade zouden kunnen 'opsparen': Meervoudige schades worden niet telkens gemeld.

Kortom, het succes van dit onderzoek valt of staat met de kwaliteit van het schaderegister. Is het register niet volledig, dan wordt de doelpopulatie (d.w.z. mensen die een schade hebben gemeld) niet goed gedekt. Daarnaast is er mogelijk vertekening door de relatieve lage respons (20%) in dit onderzoek. Dit betekent dat er een grote mate van onzekerheid is of de uitkomsten van dit onderzoek een goed beeld geven van tevredenheid van mensen met de afhandeling van aardbevingsschade. Daarnaast is van belang dat het register alleen gemelde schades bevat, en geen goed beeld geeft van alle schades. Dit is niet alleen van belang voor dit onderzoek. Zowel het eerder besproken onderzoek Gronings Perspectief (sectie 4.1) en Wonen en aardbevingen in Groningen (sectie 4.2) maken gebruik van data uit het schaderegister in de non-responscorrectie van die studies. Als ons vermoeden van vertekening in het schaderegister klopt, dan zullen de non-responscorrecties uit deze studies ook incorrect zijn. In beide studies wordt gevonden dat er in die studies relatief meer huishoudens zijn met meervoudige schade dan in het register van de NAM/het Centrum Veilig Wonen (CVW). Om die reden worden data van mensen met meervoudige schades minder belangrijk gemaakt door een kleine weegfactor toe te kennen. Wij vermoeden echter dat het schaderegister niet volledig is. Als meervoudige schades juist in het schaderegister ondervertegenwoordigd zijn, dan wordt er in Gronings Perspectief en Wonen en aardbevingen in Groningen onterecht gecorrigeerd voor het relatief hoge aantal meervoudige schades in de steekproef. Het is lastig te zeggen in hoeverre dit de conclusies van die onderzoeken zou beïnvloeden. Het is echter heel goed mogelijk dat problemen waar sommige mensen met meervoudige schades mee kampen, in werkelijkheid groter zijn dan uit de rapporten van Gronings Perspectief en Wonen en aardbevingen in Groningen blijkt.

Gezien de specifieke doelpopulatie, en het feit dat er in dit onderzoek maar weinig vragen zijn gesteld over leefbaarheid, is dit onderzoek niet erg bruikbaar voor toekomstige metingen naar leefbaarheid. Wel verdient het navolging om gebruik te maken van allerhande registers die bestaan en te maken hebben met de aardbevingsproblematiek. Informatie uit deze registers kan bijvoorbeeld worden gebruikt in non-responscorrecties.

### **Doelpopulatie**

De populatie in het onderzoek bestaat uit schademeldingen van panden die bij de NAM zijn gemeld in de periode van augustus 2012 t/m juni 2014. Dat zijn 19085 gevallen. Het totale aantal panden betreft een relatief klein deel van alle huizen die binnen het aardbevingsgebied liggen. Met andere woorden: een klein deel van de Groninger bevolking lijkt een schade te hebben gemeld. Binnen het schaderegister van de NAM en dus ook de steekproef van APE zijn schademeldingen van huurwoningen zeer ondervertegenwoordigd. Dit komt vermoedelijk omdat deze schades niet direct bij de NAM worden gemeld, maar bijna altijd via verhuurders verlopen (meestal via woningbouwcorporaties). Voor het melden van deze schades geldt een andere procedure, waardoor schades aan huurhuizen (en de effecten op de ervaren leefbaarheid onder bewoners van huurwoningen) sterk ondervertegenwoordigd zijn in het onderzoek. Wellicht zijn er bewoners geweest die wel een schade wilden melden bij de NAM, maar dat uiteindelijk niet hebben gedaan. Sommige inwoners zullen al wel schade hebben, maar die (nog) niet hebben gemeld.

Sommige schademeldingen zijn op het moment van het schrijven van het rapport in behandeling, wat het lastig maakt een categorisatie te maken naar de mate van schade. Tenslotte is onduidelijk wanneer een schade is 'gemeld'. We nemen aan dat het hier

alleen meldingen betreft die bij de juiste afdeling van de NAM terecht zijn gekomen. Sommige mensen die hun schade gemeld hebben bij een andere instantie dan NAM, zullen niet in het register staan. Al met al is het lastig de exacte kwaliteit van het register van schademeldingen dat ten tijde van het onderzoek door NAM beheerd werd, te bepalen. We durven echter wel te zeggen dat lang niet alle schades in het schaderegister staan. De doelpopulatie van het APE-onderzoek betreft mensen die schademeldingen bij de NAM hebben gedaan, en deze doelpopulatie wordt goed gedekt in dit onderzoek. Het is echter zeer de vraag of de schademeldingen in het register dekkend zijn voor alle schademeldingen, laat staan alle schadegevallen in Groningen. Met andere woorden: dit onderzoek is waardevol als het gaat om de tevredenheid van schadegevallen in het register, maar minder waardevol als het gaat om de tevredenheid met de afhandeling van schades in het algemeen. Het is wel van belang te vermelden dat het rapport van APE niet de ambitie heeft uitspraken te doen over 'alle schadegevallen', laat staan over alle inwoners van het aardbevingsgebied.

### **Steekproeftrekking**

Uit de 19000 bij de NAM gemelde schades zijn op een willekeurige manier 4250 cases geselecteerd. Omdat het vermoeden bestaat dat de ernst en type schade in sterke mate samenhangt met de ouderdom van de woning, is de steekproef in delen getrokken. Op basis van de ernst van de schade en de ouderdom (voor of na 1930) van het pand, zijn vier groepen gecreëerd (nl. nieuwe panden met veel of weinig schade en oude panden met veel of weinig schade). Voor elk van deze groepen zijn 900 panden in de steekproef terecht gekomen. De resterende 650 panden betroffen panden waar geen informatie was over mate van schade en ouderdom. Deze vorm van stratificatie zorgt ervoor dat op een efficiënte manier onderzoek kan worden gedaan naar verschillen in de uitkomstvariabelen tussen groepen respondenten. De indeling in panden met veel of weinig schade is echter niet precies. In sommige gevallen was de schade nog niet helemaal afgehandeld. Het kan dus zijn dat sommige panden onterecht zijn aangewezen als pand met een lage of hoge schade.

### **Resultaat veldwerk**

In de survey is gebruik gemaakt van een online enquête, met een herinnering. Respondenten ontvingen geen beloning voor deelname. Op verzoek konden respondenten de vragenlijst ook op papier invullen. De brieven en envelop droegen het logo van de Dialoogtafel Groningen, maar werden vanwege privacy redenen verzonden door de NAM. Dit zou een negatieve invloed kunnen hebben gehad op de hoogte van de respons.

Van de 4250 personen in de steekproef hebben 928 personen de vragenlijst ingevuld (882 online en 46 op papier). Dat is een respons van 21,8%, vergelijkbaar met Gronings Perspectief en het OTB onderzoek. Het is wel belangrijk hierbij aan te merken dat de populatie hier anders is. Het gaat om mensen die een schade hebben gemeld. Zulke mensen zouden in het algemeen toch wel iets willen zeggen over de afhandeling van die melding door de NAM, en dus zou een hogere respons te verwachten zijn. De respons is hoe dan ook laag, en er bestaat een risico dat deelnemers aan de vragenlijst verschillen van niet-deelnemers. Uit een analyse van APE bleek dat de respons representatief was met betrekking tot bouwjaar en schadebedrag. Dit is een mooi resultaat, maar het wil niet automatisch zeggen dat de respons representatief is met betrekking tot andere eigenschappen van de mensen of woningen. Het is dus maar een beperkte toets op

representativiteit, en een uitgebreidere analyse is nodig om met meer zekerheid iets te zeggen over de representativiteit van het onderzoek. Andere kenmerken van de schademelding, het pand, of de melder waren niet beschikbaar voor de onderzoekers.

### **Correctie**

Als gevolg van het gebruik van ongelijke selectiekansen en de non-respons van ongeveer 80 procent is er grote kans op vertekening in het onderzoek.

Er is voor ongelijke selectiekansen in de steekproeftrekking gecorrigeerd door middel van gewichten. Er is geen weging toegepast om te corrigeren voor de effecten van non-respons. Na communicatie via e-mail blijken de onderzoekers wel gekeken te hebben of wegingen voor non-respons zouden uitmaken voor de analyse, maar deze resultaten weken niet af, en zijn niet gerapporteerd in de onderzoeksverantwoording. De non-responsanalyses zouden eigenlijk gedaan moeten worden na correctie van ongelijke selectiekansen.

De conclusie die in het onderzoek wordt getrokken, namelijk dat er geen versturende effecten van non-respons zijn, is voorbarig. Er is op maar weinig variabelen gecorrigeerd. Er had op andere kenmerken gecontroleerd moeten worden, zeker omdat er waarschijnlijk vrij veel gegevens van de melders van de schades bekend zijn bij de NAM. Het gaat dan om allerlei kenmerken van de woningen (bijvoorbeeld: waarde, locatie, type bouw), maar ook kenmerken van de melders van de schade. Deze data waren echter voor zowel het onderzoek van APE en ook het eerder aangehaalde onderzoek van OTB Delft en Gronings Perspectief niet beschikbaar.

Omdat de respons ook gevallen bevat die nog niet geheel zijn afgehandeld, is het lastig goede correcties uit te voeren. Als er bijvoorbeeld sprake is van een samenhang tussen de duur van het hele proces van schadeafhandeling en de tevredenheid daarover, zouden de uitkomsten vertekend kunnen zijn.

### **Vragenlijst**

De vragenlijst is opgenomen in het onderzoeksrapport. In het algemeen is de kwaliteit van de vragen in orde; zo ook de uitnodigingsbrief en vraagteksten. Die zijn allemaal neutraal opgesteld en sturen de respondenten niet naar een bepaald 'goed of slecht' antwoord. De vragenlijst is vooraf getest, en is voorafgegaan door kwalitatieve interviews. Veel van de antwoordcategorieën die worden gebruikt zijn wel zogenaamde 'vague quantifiers'. Een voorbeeld hiervan is de vraag: "Hoeveel moeite kostte het om een schadeclaim in te dienen bij de NAM" met als antwoord categorieën "[Weinig], [Acceptabel] en [Veel]". Vague quantifiers worden in vragenlijstonderzoek veel gebruikt wanneer respondenten worden gevraagd naar een intensiteit van een mening of frequentie van een gedrag dat lastig in te schatten is in absolute termen bijvoorbeeld: "hoeveel moeite (...)". Hoewel veel gebruikt, zijn vague quantifiers vaak te vermijden door de vraag anders te formuleren. In het voorbeeld zou er bijvoorbeeld kunnen worden gevraagd naar hoeveel keer er contact is geweest, of hoeveel tijd de respondent heeft gependend in het proces van afhandeling van de klacht. Ook heeft een groot aantal vragen in de vragenlijst als antwoordmogelijkheden slechts 'ja' of 'nee', wat respondenten dwingt om uitgesproken antwoorden te geven. Tenslotte hebben respondenten in het onderzoek niet de mogelijkheid om aan te geven dat dat ze het antwoord op een vraag niet weten of willen geven.

## **Analyse**

Er worden in de analyse soms wel en soms geen onzekerheidsmarges vermeld. Daardoor is het niet duidelijk of geconstateerde verschillen (bijvoorbeeld in Tabel 4.5 van het rapport) 'echte' verschillen zijn, of slechts 'ruis' door toeval in de steekproef. De overige analyses worden correct uitgevoerd, hoewel de data ruimte bieden voor meer diepgaande analyses, waarin bijvoorbeeld meer ingezoomd wordt op verschillen tussen gebieden en/of tussen types schade als het gaat om de tevredenheid met de schadeafhandeling.

## **Bruikbaarheid voor effectmetingen leefbaarheid**

Het onderzoek van APE was erop gericht de tevredenheid met de schadeafhandeling van de NAM te onderzoeken. Het onderwerp van het onderzoek sluit niet goed aan bij het begrip leefbaarheid, hoewel het in het aardbevingsdossier belangrijk is te begrijpen of en in welke mate mensen tevreden zijn met overheidsinstanties en andere partijen die zich bezighouden met de aardbevingsproblematiek en de afhandeling van schades. Gezien de specifieke doelpopulatie is het onderzoek niet erg bruikbaar voor toekomstige metingen naar leefbaarheid. Wel verdient het navolging om gebruik te maken van allerhande registers die bestaan en te maken hebben met de aardbevingsproblematiek. Informatie uit deze registers kan bijvoorbeeld worden gebruikt in non-respons correcties.

## 5. Conclusies

### Goed opgezette onderzoeken

De twee doelen van onze review waren: 1) te beoordelen in hoeverre bestaande survey onderzoeken onder inwoners naar leefbaarheid kunnen dienen als basis voor het maken van beleids- en investeringsbeslissingen en 2) aanbevelingen te geven over manieren waarop toekomstig onderzoek tegemoet kan komen aan geconstateerde zwaktes en moeilijkheden van bestaand onderzoek. Deze aanbevelingen zijn erop gericht om in de toekomst goede effectmetingen te kunnen doen naar veranderingen in de leefbaarheid in het aardbevingsgebied, rekening houdend met veranderend beleid en andere maatschappelijke veranderingen (bijvoorbeeld bevolkingskrimp).

Al met al concluderen we dat alle vijf de gereviewde onderzoeken in algemene zin goed zijn opgezet. De gereviewde onderzoeken verschillen weliswaar qua inhoud, focus en methodologische opzet, maar geven gezamenlijk een goed beeld van de waargenomen leefbaarheid in Groningen en de gezondheid en veiligheidsbeleving van mensen in en buiten het aardbevingsgebied in Groningen ten tijde van de afname van de onderzoeken. Dit betekent dat de resultaten van de gereviewde onderzoeken kunnen dienen als basis voor het maken van beleids- en investeringsbeslissingen. Wanneer er in deze studies wordt gerapporteerd dat de leefbaarheid in delen van Groningen onder druk staat (bijvoorbeeld in WoON), en het hebben van aardbevingsschade gepaard gaat met een verminderde woontevredenheid of aantasting van de gezondheid (bijvoorbeeld in Gronings Perspectief), kunnen op basis van die onderzoeken beleids- en investeringsbeslissingen worden genomen. Dat zou zelfs een logische volgende stap zijn.

Dat we concluderen dat de gereviewde onderzoeken in algemene zin goed zijn opgezet en dat de resultaten van de studies bruikbaar zijn voor het nemen van beleids- en investeringsbeslissingen wil niet zeggen dat er geen enkele kanttekening te plaatsen is bij de onderzoeken. Ook zijn er specifieke sterke kanten aan de methodologische opzet van de onderzoeken die het waard zijn verder te belichten, ook met het oog op toekomstig onderzoek naar leefbaarheid en onze beantwoording van de tweede onderzoeksvraag. Dit zullen we in het vervolg van dit hoofdstuk doen. De bespreking van de sterke en minder sterke kanten van de gereviewde onderzoeken vormt de opmaat naar de volledige beantwoording van onze tweede onderzoeksvraag in [Hoofdstuk 6](#). De aanbevelingen in Hoofdstuk 6 vloeien voort uit onze conclusie in Hoofdstuk 5 en zijn specifiek gericht op toekomstig onderzoek naar (veranderingen in) leefbaarheid.

### Sterke en zwakke punten van de gereviewde onderzoeken

Het is belangrijk dat de lezer van dit rapport en al het toekomstig onderzoek naar de leefbaarheid in Groningen zich realiseert dat het heel moeilijk is 'het effect' van aardbevingen op de leefbaarheid in Groningen te bepalen (zie ook [sectie 3.4](#)). Behalve aardbevingen heeft de provincie Groningen te maken met andere trends (krimp, trek naar de stad, veranderende economie), die gelijktijdig allerlei effecten hebben op de leefbaarheid in Groningen. Wel kan worden gekeken of er binnen Groningen sterke effecten zijn door de tijd heen, en of die veranderingen in leefbaarheid door de tijd bijvoorbeeld groter of kleiner zijn binnen het aardbevingsgebied dan buiten het gebied. Of men kan Groningen als regio proberen te vergelijken met andere regio's in

Nederland. Is de trek naar de stad groter of kleiner? Is de verandering in leefbaarheid groter of kleiner?

De vijf gereviewde studies bieden hiervoor voldoende empirische data. Van belang is te vermelden dat elk van de vijf onderzoeken zijn eigen sterke kanten kent. Gronings Perspectief is breed opgezet, volgt dezelfde mensen door de tijd, kan 'causale' uitspraken doen over het effect van aardbevingen, en neemt de hele provincie Groningen mee. In algemene zin is dit het onderzoek dat het beste in staat is om uitspraken te doen over veranderingen in leefbaarheid (en andere onderwerpen die in het onderzoek centraal staan) in het aardbevingsgebied van Groningen door de tijd heen.

In de Leefbaarometer 2.0 worden niet dezelfde mensen gevolgd, en 'leefbaarheid' is vrij eng gedefinieerd. Het voordeel van dit onderzoek is wel dat dorpen, buurten en wijken binnen de provincie Groningen kunnen worden vergeleken door de tijd heen. Ook kunnen er op provincieniveau en op regionaal niveau vergelijkingen worden gemaakt met andere regio's in Nederland, en kunnen deze vergelijkingen door de tijd heen worden gedaan. Het grote nadeel van deze studie is dat 'leefbaarheid' vooral wordt gemeten met behulp van fysieke indicatoren, zoals de nabijheid van een supermarkt, of de mate van criminaliteit. Er worden slechts drie subjectieve indicatoren gebruikt die afkomstig zijn uit een survey. In die zin is deze studie wel geschikt voor een grove vergelijking van de leefbaarheid in (delen van) Groningen met andere regio's in Nederland, maar is deze niet geschikt voor het begrijpen van veranderingen in ervaren leefbaarheid in het aardbevingsgebied.

Het onderzoek naar Bevolkingskrimp en leefbaarheid dat is uitgevoerd met behulp van Groninger Panel kent twee specifieke sterke punten: Het heeft als groot voordeel dat de meeste data niet alleen de hele provincie Groningen dekken, maar ook kunnen worden vergeleken met Friesland en Drenthe. Een tweede sterk punt van dit onderzoek is dat er een uitgebreide vragenlijst rond het thema leefbaarheid wordt gebruikt. Een nadeel van het Groninger Panel is dat er (nog) geen goede longitudinale analyses worden gedaan.

Het onderzoek van OTB en CMO STAMM naar de Groninger woningmarkt en leefbaarheid kent van alle onderzoeken de engste populatie: Alleen inwoners uit negen aardbevingsgemeenten zijn in detail onderzocht. Het voordeel van deze focus op een klein gebied is wel dat er in meer detail uitspraken kunnen worden gedaan over kleine gebieden. In welke kernen is de afname van leefbaarheid sterk en waar is dat niet zo? Wat zijn hier verklaringen voor? Hoewel de totale steekproefgrootte van dit onderzoek niet heel groot was (nl. 4100 respondenten), is dit het enige onderzoek dat op viercijferig postcodeniveau met enige precisie uitspraken zou kunnen doen. Dit komt omdat dit onderzoek zich alleen richt op het aardbevingsgebied, waardoor er per kern voldoende respondenten zijn om met enige precisie uitspraken te kunnen doen. Desondanks hebben de onderzoekers ervoor gekozen ook hier niet te veel in te zoomen, en gaat het rapport dat ze hebben geschreven over 'het aardbevingsgebied' als geheel.

Het laatste rapport, van APE, over schadeafhandeling van de NAM is gebaseerd op de schademeldingen die daar zijn binnengekomen. Dit is tegelijk de sterke en zwakke kant van het onderzoek. In potentie kunnen er over de schadegevallen gedetailleerde uitspraken worden gedaan, omdat er veel informatie over elk geval beschikbaar is. De grote vraag is of de gemelde schadegevallen iets zeggen over 'de Groningers

met schade'. Ons voorzichtige antwoord op deze vraag is dat dat waarschijnlijk niet zo is, omdat er grote vraagtekens kunnen worden geplaatst bij de kwaliteit van het schaderegister. Bovendien is dit onderzoek niet longitudinaal uitgevoerd.

Na het bespreken van de sterke en minder sterke punten van ieder onderzoek willen we de rest van dit hoofdstuk gebruiken voor een wat meer algemene reflectie op de methodologie van de vijf gereviewde onderzoeken, als voorbode op de aanbevelingen die we in Hoofdstuk 6 zullen doen.

### Verschillende doelpopulaties

De onderzoeken hebben niet allemaal dezelfde doelpopulatie (zie Tabel 2). Dat maakt de onderzoeken onderling lastig vergelijkbaar. Soms wordt er onderzoek gedaan in een klein gebied (bijvoorbeeld in de negen aardbevingsgemeentes), soms in een groter gebied. Een groter gebied heeft voordelen voor vergelijkbaarheid, en de effecten van aardbevingen gelden op zich voor een groter gebied dan alleen het gebied waar bevingen plaatsvinden. Verhuizingen uit het aardbevingsgebied hebben bijvoorbeeld ook een effect op gemeenten waar naartoe wordt verhuisd. Hoe groter het gebied van de doelpopulatie echter is, hoe kleiner de steekproefgrootte per postcode-eenheid, of hoe duurder het onderzoek. Er moet een balans worden gevonden tussen vergelijkbaarheid (nl. een groot geografisch gebied) en precisie (nl. een klein geografisch gebied). Een aantal onderzoekers heeft geprobeerd om beide doelen te verenigen door dichterbij de kern van het aardbevingsgebied meer mensen te selecteren voor het onderzoek dan verder van de kern of zelfs buiten het aardbevingsgebied (d.w.z. stratificatie). Dit is een benadering die op grotere schaal navolging verdient.

Ook is onduidelijk of onderzoek zou moeten worden gedaan op het niveau van personen of huishoudens. Als het onderwerp leefbaarheid is, dan ligt een personensteekproef voor de hand (dus eventueel meerdere personen per huishouden/adres). Als het gaat om schademeldingen, dan ligt het meer voor de hand om het huishouden als analyse-eenheid te gebruiken. De onderzoeken verschillen ook hier als het gaat om de doelpopulatie.

**Tabel 2. Doelpopulatie van de vijf onderzoeken**

Onderzoek	Doelpopulatie
Gronings Perspectief (RUG)	De inwoners van de 23 Groningse gemeenten van 16 jaar en ouder
Wonen en aardbevingen in Groningen (OTB Delft)	De inwoners van de 9 aardbevingsgemeenten: Delfzijl, Eemsum, Appingedam, Loppersum, Bedum, De Marne, Slochteren, Ten Boer en Winsum
Leefbaarometer 2.0 (hierin: WoON)	Personen van 18 jaar en ouder in particuliere huishoudens in Nederland
Leefbaarheid en bevolkingskrimp in Groningen (Sociaal Planbureau Groningen)	De bevolking van de provincie Groningen van 18 jaar en ouder
Onderzoek naar de tevredenheid met de schadeafhandeling door de NAM (APE Public Economics)	Schademeldingen bij de NAM van augustus 2012 t/m juni 2014

### Hoge non-respons

De responspercentages zijn in alle onderzoeken niet erg hoog (zie Tabel 3). Dat is vervelend, want een lage respons kan leiden tot vertekeningen in de uitkomsten. De percentages rond de 20% zijn laag. Het onderwerp leefbaarheid leeft onder de Groningse bevolking, wat zou moeten resulteren in een hogere respons. We concluderen op basis van onze review en gesprekken met de betrokken onderzoekers dat de lage responspercentages waarschijnlijk deels het gevolg zijn van beperkingen in tijd (resultaten snel op dienen te leveren) en te beperkte onderzoeksbudgetten. Tegelijk gaven onderzoekers en leden van de stuurgroep aan dat er ook sprake van onderzoeksmoeheid onder inwoners in Groningen. Die onderzoeksmoeheid zou een rol kunnen spelen bij de hoge non-respons in de gereviewde onderzoeken.

Het is lastig te zeggen hoe hoog de respons precies zou moeten zijn in de gereviewde onderzoeken en in toekomstig onderzoek in Groningen. Zelfs bij hoge respons bestaat er namelijk een kans op vertekening door non-respons. We vinden het echter wel wenselijk dat de respons op onderzoeken hoger ligt dan in de gereviewde onderzoeken en denken ook dat dit ook haalbaar is wanneer er voldoende tijd en middelen worden besteed aan het uitvoeren van het veldwerk. We denken dat—ook bij onderzoeksmoeheid, een respons van 40-60 procent haalbaar is in toekomstig leefbaarheidsonderzoek onder inwoners in Groningen. Ter illustratie, in online onderzoek binnen het LISS panel (Langlopende Internet Studies voor Sociale wetenschappen) dat bestaat uit huishoudens uit alle lagen van de Nederlandse bevolking werd in 2008 een responspercentage van 62% bereikt op de eerste vragenlijst, en een deelnamepercentage aan het panel van 48%. In Hoofdstuk 6 komen we met aanbevelingen gericht op het realiseren van hogere responspercentages.

**Tabel 3. Percentage respons in de vijf onderzoeken**

Onderzoek	Percentage respons
Gronings Perspectief (RUG)	15,7% (meting 1 binnen panel)
Wonen en aardbevingen in Groningen (OTB Delft)	20,8%
Leefbaarometer 2.0 (hierin WoON)	57,0%
Leefbaarheid en bevolkingskrimp in Groningen (Sociaal Planbureau Groningen)	11,0% (vier jaar na start van panel)
Onderzoek naar de tevredenheid met de schadeafhandeling door de NAM (APE Public Economics)	21,8%

### Geen onzekerheidsmarges

Alle vijf onderzoeken zijn gebaseerd op steekproeven. Dat brengt met zich mee dat er een element van onzekerheid zit in de uitkomsten. Geconstateerde verschillen kunnen 'echte' verschillen zijn, maar ook 'ruis' van de steekproef. Om te voorkomen dat artefacten van de steekproef aangezien worden voor echte effecten, moeten de onzekerheidsmarges worden vermeld. Dit gebeurt alleen in Gronings Perspectief en het WoON-onderzoek systematisch. In de andere onderzoeken worden onzekerheidsmarges niet systematisch of helemaal niet vermeld. Daardoor krijgen de uitkomsten ook nog eens een schijn van zekerheid die er niet is.

### **Weegprocedures kunnen verbeterd worden**

Alle onderzoeken hebben last van hoge non-respons. Geen van de onderzoeken rapporteert uitgebreid over non-responsconversie, of non-responsanalyses. Dat is jammer, maar wellicht reikten de beschikbare middelen niet zover. Daarnaast wordt er vaak gebruik gemaakt van complexe steekproefdesigns, waarin sommige categorieën mensen een hogere kans krijgen om in de steekproef te komen dan andere. In deze gevallen is het noodzakelijk om de resultaten van het onderzoek te corrigeren. De weegmethodes die worden gebruikt in de gereviewde onderzoeken zijn technisch gezien soms niet correct, hoewel het gebruik van een betere methode maar heel kleine effecten zal hebben op de uitkomsten. Belangrijker is dat er maar weinig variabelen worden gebruikt om te wegen, en dat die variabelen maar beperkt samenhangen met de uitkomstvariabelen. Voor een goede weegprocedure moet specifieke informatie worden gevonden voor de hele populatie in registers. Het schaderegister van de NAM/ het Centrum voor Veilig Wonen biedt daarvoor een goede mogelijkheid, die door een aantal studies ook wordt benut. Er bestaan echter twijfels over de kwaliteit van dit register.

Meer en andere weegvariabelen zouden beschikbaar zijn in registers van de overheid, denk hierbij bijvoorbeeld gegevens van het CBS (per postcodegebied), data van de provincie, gemeentes, het kadaster of de GGD.

### **Bijna geen echte longitudinale analyses**

In Gronings Perspectief, en wellicht in toekomstig onderzoek van het Sociaal Planbureau Groningen, worden longitudinale metingen gedaan. Longitudinale metingen—waarbij dezelfde mensen door de tijd heen worden gevolgd—bieden kansen om bijvoorbeeld na te gaan welke respondenten zich door de tijd heen veiliger of juist onveiliger gaan voelen, en te proberen veranderingen in ervaren leefbaarheid te verklaren. Deze mogelijkheden worden momenteel nog onderbenut. Deels kunnen deze analyses worden gedaan op de bestaande data. Voor een goede monitor en goede toekomstige effectmetingen rondom het thema leefbaarheid is het echter noodzakelijk dat toekomstig onderzoek echt longitudinaal wordt opgezet, en dat er in de opzet van het onderzoek rekening wordt gehouden met het terugkerend karakter van de metingen. Dat houdt in dat er systematisch wordt nagedacht over welke vragen wel en niet terugkerend moeten worden gevraagd, hoe uitval van respondenten te voorkomen, en hoe daar op een goede manier voor te compenseren.

### **Data zijn onder-geanalyseerd**

Er is de afgelopen jaren een relevante hoeveelheid data verzameld rond het thema leefbaarheid. Hoewel de onderzoeken van elkaar verschillen in opzet en gestelde vragen, zijn er ook veel overeenkomsten. Die liggen met name in de kenmerken van de woningen en respondenten die worden gemeten. In alle onderzoeken wordt gevraagd naar leeftijd, geslacht, opleiding, woonplaats, afkomst, en in vrijwel alle onderzoeken naar inkomen, het hebben van schade, type schade en of en hoe die afgehandeld is. Een uitsplitsing van de resultaten op de belangrijkste uitkomstvariabelen naar deze achtergrondkenmerken is zinvol. Het maakt dat de resultaten van de onderzoeken beter te vergelijken en te integreren zijn. Dit is waardevol voor onderzoekers (biedt mogelijkheden voor cross-validatie) en voor beleidsmakers (helpt hen bij duiden van de resultaten en vervolgens ook bij het nemen van beleids- en investeringsbeslissingen). Deze uitsplitsing vindt nog onvoldoende plaats in de gereviewde studies. Dit maakt dat

we vinden dat in algemene zin de data onder-geanalyseerd zijn. Dit wil zeggen dat uit de gereviewde studies er nog 'meer' te halen is en additionele vragen beantwoord kunnen worden door extra analyses te doen. Bestaande vragen die leven bij beleidsmakers zouden wellicht goed met een of meerdere van de door ons gereviewde datasets kunnen worden beantwoord.

### **Het meten van leefbaarheid**

Leefbaarheid is een breed begrip. Inwoners van het aardbevingsgebied zullen vaak anders denken over leefbaarheid, en welke aspecten daarin voor hen van belang zijn dan bewoners buiten Groningen. Het is in die zin bijna onmogelijk om leefbaarheid uitputtend te meten met een vragenlijst. De vijf onderzoeken verschillen echter behoorlijk in de manier waarop zij leefbaarheid proberen te meten. Dat komt omdat de onderzoeken allemaal een andere focus hebben. Gronings Perspectief richt zich met name op mentale gezondheid, 'Wonen en aardbevingen' op de woningmarkt, het onderzoek van APE op de tevredenheid met schadeafhandeling. Alleen het onderzoek van het Sociaal Planbureau Groningen richt zich expliciet op het meten van 'leefbaarheid'. Er zijn maar weinig vragen die in bijna alle onderzoeken worden gesteld. De belangrijkste daarvan zijn de drie subjectieve vragen uit de WoON-survey die worden gebruikt in de Leefbaarometer. Kortom, de resultaten uit de verschillende onderzoeken zijn lastig een op een vergelijkbaar, omdat de vragenlijsten weinig overeenkomen.

## 6. Aanbevelingen

De goede lezer heeft wellicht aan de hand van de conclusies in [Hoofdstuk 5](#) al een aantal aanbevelingen kunnen afleiden die zich richten op de methodologische opzet van toekomstige studies. De vijf gereviewde studies vormen een goede basis voor kennis over de leefbaarheid in en rond het aardbevingsgebied in Groningen, maar zijn in de huidige opzet en ook gezien de (vooralsnog) onzekere looptijd en bijbehorende financiering nog niet geschikt om de komende jaren goed te meten hoe leefbaarheid verandert, laat staan om mogelijke effecten van veranderingen in beleid te monitoren. Om in de toekomst effectmetingen te kunnen doen naar veranderingen in de leefbaarheid in het aardbevingsgebied, rekening houdend met veranderend beleid en andere maatschappelijke veranderingen (bijvoorbeeld bevolkingskrimp) is het noodzakelijk dat:

### **Aanbeveling 1: Er moet worden geïnvesteerd in longitudinaal onderzoek. Dat wil zeggen: onderzoek waar gedurende een langere periode dezelfde mensen worden ondervraagd.**

Het is wenselijk dit onderzoek langdurig op te zetten. Dat wil zeggen dat voor een periode van 5-10 jaar eenzelfde onderzoeksdesign wordt gebruikt, en dat op regelmatige basis metingen plaatsvinden om veranderingen op een eenduidige manier te kunnen meten. Het is hierbij belangrijk dat een bepaalde set van vragen op identieke wijze periodiek herhaald wordt. Zulk onderzoek wordt nu al uitgevoerd in het Groninger Panel en het Gronings Perspectief. Hierbij kan gedacht worden om bijvoorbeeld de vragenlijsten altijd te beginnen met twee vaste blokken vragen. Daarnaast kunnen er aanvullend nog vragenblokken worden toegevoegd die per meting verschillen en die de benodigde flexibiliteit waarborgen die nodig is voor langlopend onderzoek naar leefbaarheid in Groningen. Hierdoor kan er bijvoorbeeld ingesprongen worden op veranderend beleid als het gaat om gaswinning en de afhandeling van schades, maar ook op overige veranderende economische en politieke omstandigheden. Als het doel van toekomstig onderzoek is om effectmetingen te doen, dan is het naast een longitudinaal karakter van de data namelijk ook van belang om alternatieve verklaringen voor veranderingen in de leefbaarheid zoveel mogelijk uit te sluiten. Dat kan door het meten van zulke verklaringen in de vragenlijst, maar ook door het onderzoek zó op te zetten dat er een of meer goede vergelijkingsgroepen zijn.

Een tweede reden om meer te investeren in onderzoek naar leefbaarheid betreft het feit dat veel van de gereviewde data ondergeanalyseerd zijn. Dat wil zeggen dat de data ruimte bieden voor extra analyses over bijvoorbeeld ruimtelijke verschillen of individuele verschillen tussen huishoudens. Die analyses kosten extra tijd en dus geld. Het kan voor individuele gemeenten en/of voor de provincie zinvol zijn om in toekomstige WoON-onderzoeken de steekproef te vergroten zodat analyses over de leefbaarheid kunnen worden gedaan op gemeentelijk of dorpsniveau in plaats van op regionaal niveau.

In persoonlijke gesprekken met onderzoekers gaven zij aan dat de huidige budgetten voor onderzoek naar leefbaarheid en aanverwante onderwerpen niet toereikend zijn om bijvoorbeeld het veldwerk optimaal uit te voeren (bijvoorbeeld qua respons, non-responsconversie), of alle data uitputtend te analyseren. Bovendien zijn de financieringsmogelijkheden voor onderzoek naar de bovengrondse non-technische gevolgen van gaswinning beperkt. Hoewel wij als onderzoekers geen inzicht hebben in

de exacte begrotingen van elk onderzoek, onderschrijven we dat goed (longitudinaal) onderzoek niet goedkoop is en er budget voor vrijgemaakt moet worden. Onderzoek naar de sociale consequenties van de aardbevingsproblematiek in Groningen is belangrijk, niet alleen voor het leren begrijpen van de gevolgen van aardbevingen voor de inwoners van Groningen, maar ook voor het opstellen van toekomstig beleid dat de leefbaarheid in het aardbevingsgebied borgt en bevordert. Er gaat (terecht) veel geld naar technisch onderzoek naar de gevolgen van aardgaswinning voor de ondergrond; maar er dient ook geld vrij gemaakt worden voor sociaalwetenschappelijk onderzoek naar de gevolgen voor de bovengrond. Als de leefbaarheid in Groningen en het welzijn en gezondheid van inwoners centraal staan, dan dient dit ook tot uitdrukking te komen in de financieringsmogelijkheden en middelen voor sociaalwetenschappelijk onderzoek.

In het opzetten van zulk longitudinaal onderzoek is het van belang dat helder is wat er precies gemeten moet worden. Dit betekent ook dat er voldoende tijd en middelen moeten worden uitgetrokken aan de 'voorkant' van het onderzoek; er dient veel overlegd te worden tussen de opdrachtgever en de onderzoekers. Vragen die bijvoorbeeld belangrijk zijn om te beantwoorden, zijn onder andere: Welke aspecten van leefbaarheid zijn belangrijk om te monitoren en waarom? Welke beleidsvraag dient het onderzoek te beantwoorden? Op welk geografisch niveau moeten de analyses worden uitgevoerd? Wat zouden goede vergelijkingsgroepen kunnen zijn, en hoe kunnen die vergelijkingen zo goed mogelijk worden gemaakt? Dit leidt tot onze tweede aanbeveling:

**Aanbeveling 2: Het is van belang dat in het uitzetten en begeleiden van onderzoek voldoende inhoudelijke en methodologische expertise aanwezig is om de kwaliteit van het longitudinale onderzoek te kunnen waarborgen.**

Om effectmetingen te doen, moeten in longitudinaal onderzoek vooraf een aantal belangrijke keuzes worden gemaakt, die te maken hebben met de inhoud van het onderzoek (bijvoorbeeld: te meten constructen en daaruit voortkomende enquêtevragen), de manier waarop mensen worden geworven, en de manier waarop analyses worden gedaan. De opdrachtgever en onderzoekers maken deze keuzes gezamenlijk, bij voorkeur ook na consultatie van een begeleidingsgroep. Het is hierbij belangrijk dat relevante partijen voldoende inhoudelijke en methodologische expertise hebben, en zo geïnformeerd keuzes maken om de kwaliteit van het onderzoek te waarborgen. De volgende aanbevelingen zoomen specifiek in op keuzes in het design van longitudinale surveys.

**Aanbeveling 3: We raden aan om in toekomstig onderzoek ervoor te kiezen een complex steekproefdesign te gebruiken.**

We denken dat het zinvol is om alle inwoners van Groningen te betrekken bij het onderzoek, en wellicht ook inwoners uit de andere noordelijke provincies. Een onderzoek in een groot geografisch gebied heeft echter als nadeel dat de aantallen respondenten in individuele dorpen en gemeenten soms klein zijn, tenzij er een enorme steekproef wordt getrokken. We pleiten er daarom voor om inwoners van 'het aardbevingsgebied' een veel hogere kans te geven in de steekproef te komen (dat wil zeggen, de steekproef gestratificeerd te trekken). Op die manier kan er in het aardbevingsgebied op detailniveau uitspraken worden gedaan (bijvoorbeeld uitgesplitst per gemeente of woonkern), en kan tegelijkertijd het aardbevingsgebied worden vergeleken met andere gebieden binnen, of eventueel buiten Groningen. Er kan ook voor gekozen worden om dit design nog verder

door te voeren, en ervoor te kiezen om binnen het aardbevingsgebied een dorp of gebied te kiezen met een nog hogere kans. Deze laatste benadering is bijzonder zinvol wanneer er wordt gekozen voor een mixed-methods benadering (zie aanbeveling 10).

In veel van de door ons gereviewde studies was de non-respons hoog. Het halen van een hoge respons, en het analyseren van selectiviteit in die non-respons is van groot belang voor de generaliseerbaarheid van de uitkomsten van het onderzoek.

#### **Aanbeveling 4: In toekomstig onderzoek moet een specifieke strategie worden gebruikt om non-respons te verkleinen, met name bij moeilijk bereikbare groepen.**

Op zich is het verhogen van de respons alleen niet voldoende om te garanderen dat de resultaten uit een survey representatief zijn voor een populatie. De vertekening wordt behalve door de hoogte van de respons ook bepaald door de variatie in het responsgedrag. Naarmate de variatie in de responskansen groter is, zal de vertekening ook groter zijn.

Het is belangrijk dat onderzoekers zich realiseren dat er altijd non-respons zal optreden in hun surveys. Bij het opzetten van een survey moeten ze daarmee al rekening houden. Dat dient op twee manieren te gebeuren:

1. Het veldwerk dient erop gericht te zijn het percentage non-respons zo laag mogelijk te houden. Dat kan op allerlei manieren, bijvoorbeeld door middel van gedegen training van interviewers, het aanbieden van beloningen en het sturen van herinneringsboodschappen (per brief of e-mail). Het is hierbij belangrijk om voldoende tijd uit te trekken voor dataverzameling: als resultaten van een onderzoek binnen drie weken binnen moeten zijn dan is er te weinig tijd voor reminders en blijft de respons achter. Ook dient er voldoende budget beschikbaar te worden gesteld om de non-respons te verkleinen. Voor onderzoek naar leefbaarheid in Groningen kan onderzoeksmoeheid daarnaast een negatief effect hebben op de respons: mensen worden overvraagd en/of ze zijn niet meer gemotiveerd om deel te nemen aan onderzoek omdat ze zich niet gehoord voelen of hun antwoorden niet vertaald zien in beleid. Om dit (deels) te ondervangen is het belangrijk dat onderzoek naar leefbaarheid onderdeel is van een beleidscyclus (zie aanbeveling 9).
2. In de survey moeten relevante hulpvariabelen worden gemeten. Dat zijn variabelen die gemeten worden in de vragenlijst en waarvoor de verdeling in de populatie bekend is. Met die variabelen kan, in de eerste plaats, een non-responsanalyse worden uitgevoerd. Per variabele kan worden nagegaan of de verdeling in de respons afwijkt van de verdeling in de populatie. Zo wordt inzicht verkregen in de mogelijke selectiviteit van de non-respons. En als de non-respons selectief is dan kunnen de desbetreffende hulpvariabelen worden gebruikt in een weegprocedure. Hierbij dient opgemerkt te worden dat lang niet elke hulpvariabele effectief is in de correctie van selectiviteit. Effectieve weegvariabelen dienen te voldoen aan twee voorwaarden:
  - i) Ze moeten sterk gecorreleerd zijn met het responsgedrag (de responskansen). Dit kan worden vastgesteld in de non-responsanalyse.
  - ii) De hulpvariabelen moeten sterk gecorreleerd zijn met de doelvariabelen van het onderzoek (bijvoorbeeld leefbaarheid). Dit is minder eenvoudig vast te stellen. De samenhang kan worden onderzocht voor de respons, maar de

uitkomsten hoeven niet automatisch te gelden voor de non-respons. Soms bieden de uitkomsten van ander onderzoek meer inzicht.

Het is van belang dat er altijd een non-responsanalyse wordt uitgevoerd en dat daarvan verslag wordt gedaan in de onderzoeksverantwoording. Het verdient bijvoorbeeld aanbeveling om de vergelijking van de verdeling van de hulpvariabelen in de respons en in de populatie in een appendix op te nemen. Ook dient gedetailleerd te worden gedocumenteerd hoe het gebruikte weegmodel eruitziet. Tenslotte dient er een inschatting te worden gemaakt van hoe effectief de correctieprocedure was bij het verminderen van de vertekening. Dit brengt ons bij onze volgende aanbeveling:

**Aanbeveling 5: Er moet op een betere manier worden gewogen.**

Zeker wanneer toekomstig onderzoek op meerdere schaalniveaus wordt uitgevoerd (bijvoorbeeld op schaal van de provincie, het aardbevingsgebied of gebieden daarbinnen) is het noodzakelijk om op een goede manier te corrigeren voor het design van de steekproef en non-respons. Het is noodzakelijk voor de opzet van het onderzoek na te gaan met welke informatie er uit registers gewogen kan worden. Dit kan informatie zijn die beschikbaar is bij bijvoorbeeld gemeentes, het kadaster, het Centraal Bureau voor de Statistiek, Gezondheidsdiensten of de provincie, maar ook bijvoorbeeld bij schaderegisters.

**Aanbeveling 6: Een goed register van schadegevallen kan helpen bij het trekken van steekproeven of het corrigeren van non-respons.**

Het hebben van schade aan de woning is een belangrijk thema binnen het aardbevingsdossier. Gezien de omvang van de schade valt te voorzien dat het afhandelen van schadegevallen nog jaren zal duren. Daarmee heeft het direct en indirect een impact op de waargenomen leefbaarheid van buurten en regio's. Informatie over schades kan heel behulpzaam zijn bij het analyseren en corrigeren van bijvoorbeeld non-respons. Het is dan wel van belang dat dit register van goede kwaliteit is, dat alle schades gemeld en geregistreerd worden, en dat het schaderegister en onder gespecificeerde voorwaarden toegankelijk voor onderzoekers.

**Aanbeveling 7: Toekomstig onderzoek zou expliciet surveydata moeten proberen te koppelen aan andere data.**

Dit kunnen data zijn van de NAM, het Centrum voor Veilig Wonen, de NCG, maar bijvoorbeeld ook van die gemeentes en andere overheden. Deze aanbeveling sluit aan bij aanbeveling 5, maar het doel is hier anders. Geaggregeerde registerdata kan worden gebruikt bij het verbeteren van non-responsanalyses en correctieprocedures. Data op individueel niveau bieden kansen voor het doen van meer diepgaande analyses op het individuele niveau van respondenten. Gronings Perspectief is het enige onderzoek waarin expliciet wordt getracht onderzoeksresultaten aan elkaar te knopen en vergelijken. Dat betreft het onderzoek van Lifelines. In dit onderzoek worden veel multivariate en longitudinale modellen geschat, die recht doen aan de rijke data die deze studie oplevert.

### **Aanbeveling 8: De inhoud van de vragenlijsten moet worden afgestemd tussen en binnen de verschillende onderzoeken.**

Hierbij zien wij een rol voor opdrachtgevers, beleidsmakers, inhoudelijk deskundigen op het gebied van leefbaarheid, en onderzoekers. Als het gaat om de inhoud van de onderzoeken, dan is over het algemeen de kwaliteit van de vragenlijsten in bestaande studies in orde. Er komen wel veel verschillende thema's aan bod in de verschillende vragenlijsten. Elk onderzoek gebruikt andere vragenlijsten en heeft andere onderzoeksvragen. Dit zorgt tot een heel gefragmenteerd beeld van wat leefbaarheid inhoudt. Hoewel in sommige gevallen de vragenlijst wel getest is, zou het ontwikkelen en testen van de vragenlijsten in de verschillende onderzoeken meer aandacht en kwalitatieve interviews behoeven.

### **Aanbeveling 9: Onderzoek naar leefbaarheid dient onderdeel te zijn van een beleidscyclus.**

In deze aanbeveling komen de eerdere aanbevelingen deels samen. Voor toekomstige onderzoek naar leefbaarheid in Groningen en de daaruit voortvloeiende beleids- en investeringsbeslissingen is het essentieel dat onderzoek naar leefbaarheid onderdeel wordt van een beleidscyclus. Dit borgt niet alleen de relevantie en kwaliteit van onderzoek naar leefbaarheid, maar geeft inwoners van Groningen ook het gevoel dat ze gehoord worden en dat onderzoeksresultaten omgezet worden in relevant beleid. De NCG zou dit ons inziens kunnen realiseren. De NCG zou dit bijvoorbeeld kunnen doen door onderzoeken naar leefbaarheid onder te brengen bij en te koppelen aan het NCG Meerjarenplan Aardbevingsbestendig en Kansrijk Groningen. De betrokken partijen zouden gezamenlijk de algemene onderzoeksvragen, doelvariabelen, onderzoeksvragen en onderzoeksmethodiek plus analysemethoden en wensen kunnen bepalen. De uitkomsten van de leefbaarheidsonderzoeken worden vervolgens actief gecommuniceerd naar inwoners van Groningen. Ook in openbare beleidsstukken zoals het Meerjarenplan zouden de resultaten van leefbaarheidsonderzoeken terug moeten komen en expliciet gekoppeld moeten worden aan beleids- en investeringsbeslissingen. Inwoners van Groningen voelen zich zo gehoord en kunnen er vertrouwen in hebben dat er iets gedaan wordt met hun antwoorden. Dit is niet alleen belangrijk voor het verhogen van de responspercentages van de onderzoeken, maar ook voor behoud en herstel van vertrouwen van Groningers in overheden.

Tot slot nog een aanbeveling in de categorie 'overig'. Dit is een aanbeveling die niet direct aansluit bij het tweede doel van ons onderzoek—het doen van aanbevelingen voor toekomstige effectmetingen—, maar betreft een aanbeveling die we naar aanleiding van onze ervaringen bij het schrijven van deze review hebben opgedaan.

Hoewel deze review zich heeft gericht op surveys, hebben we in het persoonlijk contact met onderzoekers vaak gehoord dat juist de kwalitatieve component van die onderzoeken bijzonder zinvol was. Het gaat hier bijvoorbeeld om versnellingskamers (zoals in het OTB-onderzoek), focusgroepen, en interviews (zoals in het Gronings Perspectief). Wij willen deze suggestie graag onderschrijven.

**Aanbeveling 10: Mixed-methods onderzoek, waar kwalitatieve gegevens worden gecombineerd met surveys, biedt een meerwaarde bij het begrijpen van een begrip als leefbaarheid.**

In het kader van een longitudinaal onderzoek zouden inwoners van een of meerdere kernen bijvoorbeeld kunnen worden gevolgd door middel van een survey, en kan een aantal daarnaast (herhaaldelijk) worden geïnterviewd.

## 7.

## Referentielijst

- Baker, R., Blumberg, S. J., Brick, J. M., Couper, M. P., Courtright, M., Dennis, J. M., ... Zahs, D. (2010). Research synthesis: AAPOR report on online panels. *Public Opinion Quarterly*, 74, 711–781.
- Bethlehem, J. G. (2009). *Applied survey methods – A statistical perspective*. Hoboken, NJ, VS: John Wiley & Sons.
- Bethlehem, J. G. (2012). *Peilingen beoordelen, een checklist*. Verkregen op: <https://www.cbs.nl/nl-nl/publicatie/2012/34/peilingen-beoordelen>
- Bethlehem, J. G., & Biffignandi, S. (2012). *Handbook of web surveys*. Hoboken, NJ, VS: John Wiley & Sons.
- Bethlehem, J. G., Cobben, F., & Schouten, B. (2011). *Handbook of non-response in household surveys*. Hoboken, NJ, VS: John Wiley & Sons.
- Couper, M. P. (2008). *Designing effective web surveys*. New York, NY, USA: Cambridge.
- Horvitz, D.G., & Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47, 663–685.
- Knol, F., Maas, C., & Roes, T. (2002). *Zekere banden, sociale cohesie, leefbaarheid en veiligheid*. De Hart, J. (Red.). Den Haag: Sociaal Cultureel Planbureau.
- Krug, S. (2006). *Don't make me think! A common sense approach to web usability* (2<sup>e</sup> ed.). Berkeley, CA, USA: New Riders.

# Bijlage 1: Offerte aanvraag (uitnodiging) NCG/EZ



Ministerie van Economische Zaken

> Retouradres Postbus 93144, 2509 AC Den Haag

Universiteit Leiden  
mevrouw E. ter Mors  
emors@fsw.leidenuniv.nl

**Inkoop Uitvoering Centrum EZ**

**Bezoekadres**  
Prinses Beatrixlaan 2  
2595 AL Den Haag

**Postadres**  
Postbus 93144  
2509 AC Den Haag  
[www.rvo.nl](http://www.rvo.nl)

**Behandeld door**  
G. Strootman

Datum 22-9-2016  
Betreft Offerteaanvraag "review op uitgevoerde surveys"

**Ons kenmerk**  
20161039

**Bijlagen**  
2

Geachte mevrouw ter Mors,

Hierbij nodig ik u uit om namens de Nationaal Coördinator Groningen een offerte uit te brengen voor het uitvoeren van een review op uitgevoerde surveys binnen het aardbevingsdossier van Groningen.

**Beschrijving van de opdracht**

Binnen het aardbevingendossier zijn de afgelopen jaren vele survey studies onder de bevolking uitgevoerd over uiteenlopende onderwerpen. Het gaat bijvoorbeeld om studies die (on)tevredenheid over de afhandeling van schades in kaart brengen en studies die tevredenheid met de woonomgeving en factoren die hierbij bepalend zijn onderzoeken. Diverse partijen, waaronder NCG en NAM, willen (deels) op basis van deze studies (beleids- en investerings-) beslissingen nemen die bijdragen aan het verbeteren van de veiligheid en leefbaarheid van Groningen. De verscheidenheid aan beschikbare surveys studies—verscheidenheid zowel qua opzet als focus en resultaten—maakt het echter lastig om de toepasbaarheid van de resultaten van de verschillende studies te beoordelen.

**Nadere toelichting van de opdracht**

Wat kun je op basis van een bepaalde studie en bijbehorende steekproef bijvoorbeeld concluderen en wat niet? Is een steekproef representatief of niet, hoe zit het met nonrespons? Heeft het moment van ondervragen mogelijk impact gehad op de resultaten? Wat is de kwaliteit van de vragen, in welke mate zijn deze sturend en is er ook de mogelijkheid om geen mening te hebben en geven? Kloppen de analyses en bijbehorende conclusies?

Een manier om deze en andere vragen te beantwoorden is door een aantal bestaande surveys te laten reviewen door een team van wetenschappers. In eerste instantie zou dit bijvoorbeeld om 5 surveys kunnen gaan. Niet alleen geeft het reviewen van bestaande surveys inzicht in de bruikbaarheid bevindingen, ook bieden de 'do's en don'ts die blijken uit de reviews (i.e., sterke en zwakke punten van studies zoals geïdentificeerd door de reviewers) handvatten om de kwaliteit

van toekomstig survey onderzoek te borgen en verbeteren. Een dergelijk wetenschappelijk reviewteam zou kunnen bestaan uit 3-5 wetenschappers met expertise op het gebied van methodologie en kwantitatief sociaalwetenschappelijk onderzoek.

De opdracht die wordt verstrekt gaat om een combinatie van zowel methodologische expertise op het gebied van eenmalige als longitudinale kwantitatieve surveys (o.a. expertise op het gebied van steekproeftrekking, representativiteit, weging, respons, non-respons, vraagstelling, antwoordmogelijkheden/format, analyses/conclusies, etc.). Daarnaast beschikt men over sociaalwetenschappelijke expertise op het gebied van publieke perceptie rondom (ongewenste) energieprojecten en destructieve rechtvaardigheid, compensatie, etc.: hoe dit te onderzoeken). Hierbij dient te worden samengewerkt, overleg worden gepleegd met stakeholders en vindt regelmatig overleg plaats met de stuurgroep.

De opdracht bestaat uit 2 fasen:

Fase 1: Oriëntatie, afbakening opdracht, bepalen werkwijze :

- Gezamenlijk aanscherpen, afstemmen en afbakenen van de opdracht. Wetenschappers en stuurgroep bepalen welke 5 surveys gereviewd worden, de focus en vorm van het review (e.g., kwaliteit steekproef, selection/participation bias, non-respons, kwaliteit van vragen/analyses, conclusies) en ambitieniveau van het review (e.g., baseren op rapporten alleen en/of bij ontbrekende informatie contact zoeken met betrokken onderzoekers: dit laatste is waarschijnlijk nodig).

Fase 2: Review/opleverfase

Wetenschappers:

- Het uitvoeren van de 5 reviews, parallel of sequentieel (zoals bepaald in fase 1).
- Het opschrijven van bevindingen: een afzonderlijk document voor elke gereviewde survey
- Een notitie met waarin de conclusies/leerpunten/aanbevelingen voor vervolgonderzoek en toekomstige surveys staan.

Overig tussentijds overleg (meetings, telecons) tussen wetenschappers onderling en wetenschappers en stuurgroep om de voortgang, inhoud/focus van het onderzoek en de resulterende rapporten te bespreken.

De doorlooptijd van dit onderzoek wordt geschat op ongeveer 5 maanden.

#### **Uitvoeringsperiode**

De opdracht zal aanvangen bij verstrekking van de opdrachtbrief aan de opdrachtnemer en eindigt van rechtswege na uitvoering van alle diensten en betaling door het NCG. Het definitieve eindproduct en de hieraan gerelateerde diensten dienen na 5 maanden opgeleverd te zijn, tenzij partijen een herziene opleverdatum schriftelijk zijn overeengekomen.

#### **Communicatie**

Alle communicatie met betrekking tot dit offertetraject dient te verlopen via Gea Strootman, IUCEZteam8@rvo.nl, inkoop- en contractmanager van het Inkoop

Uitvoering Centrum. Uw vragen of opmerkingen kunt u ook aan deze persoon richten.

Als u een klacht over dit offertetraject heeft, dan kunt de klachtenprocedure opvragen bij: [KlachtenmeldpuntAanbesteden@minez.nl](mailto:KlachtenmeldpuntAanbesteden@minez.nl). Uw klacht kunt u conform de in die procedure vermelde wijze, ook via genoemd mailadres indienen.

### **Eisen**

Naast de hierboven gestelde eisen, worden aan de werkzaamheden de volgende eisen gesteld.

- De in te zetten wetenschappers hebben methodologische expertise op het gebied van zowel eenmalige als longitudinale kwantitatieve surveys (o.a., expertise op het gebied van steekproeftrekking, representativiteit, weging, respons, nonrespons, vraagstelling, antwoordmogelijkheden/format, analyses/conclusies, etc.) en sociaal-wetenschappelijke expertise op het gebied van publieke perceptie rondom (ongewenste) energieprojecten (acceptatie, risicoperceptie, zorgen, tevredenheid met de woonomgeving, vertrouwen, waargenomen procedurele en distributieve rechtvaardigheid, compensatie, etc.
- Het is voor het slagen van het project verder belangrijk dat de partijen in het review team een goede (onafhankelijke) reputatie hebben en geen belanghebbende zijn.
- de maximum/vaste totaalprijs is niet hoger dan €100.000,-.

Het niet voldoen aan een of meer eisen, betekent dat u offerte terzijde wordt gelegd en niet voor gunning in aanmerking komt.

### **Inhoud van de offerte**

Ik verzoek u om in de offerte de volgende onderwerpen op te nemen:

- Een beschrijving van de onderzoeksaanpak, waarbij u aangeeft welke onderzoeksmethoden u wilt toepassen, met een motivering waarom deze methoden volgens u het meest geschikt zijn.
- Een raming van het aantal uren met bijbehorende uurtarieven, per activiteit/fase/medewerker. Tevens offreert u de vaste totaalprijs (exclusief en inclusief btw) voor uitvoering van de opdracht.

### **Indiening van de offerte**

Uw schriftelijke offerte moet uiterlijk 3 oktober, 12 uur 's middags zijn ingediend.

U stuurt uw offerte naar:

IUCEZteam8@rvo.nl ter attentie van: Gea Strootman

De datum 3 oktober, 12 uur 's middags van indienen, *dient* als een fataal moment te worden beschouwd. Het risico van vertraging tijdens de verzending of door onjuiste of onvolledige adressering is voor u.

Offertes en stukken die niet op genoemde datum en tijdstip in het bezit van IUCEZ zijn, worden van de beoordeling van de offertes uitgesloten.

### **Beoordeling van de offerte**

De opdracht zal worden gegund aan de economisch meest voordelige offerte.

Beoordeling vindt plaats op:

- Bekendheid met dit specifieke gaswinning/aardbevingen dossier en ontwikkelingen in de regio.
- Dat partijen een goed netwerk hebben.
- Ervaring met het werken binnen gevoelige dossiers.

- Goed kunnen samenwerken, overleggen, communiceren met een (stuur)groep met een diverse achtergrond.
- Een beschrijving van de onderzoeks aanpak, waarbij u aangeeft welke onderzoeksmethoden u wilt toepassen, met een motivering waarom deze methoden volgens u het meest geschikt zijn.
- Een raming van het aantal uren met bijbehorende uurtarieven, per activiteit/fase/medewerker. Tevens offreert u de vaste totaalprijs (exclusief en inclusief btw) voor uitvoering van de opdracht.

Ik verwacht omstreeks 10 oktober 2016 te beslissen over het verlenen van de opdracht.

**Van toepassing zijnde voorwaarden**

Op de te verlenen opdracht zijn de Algemene Rijksvoorwaarden voor het verstrekken van opdrachten tot het verrichten van Diensten 2014 (ARVODI 2014), zie bijlage 'Algemene Rijksvoorwaarden voor het verstrekken van opdrachten tot het verrichten van diensten 2014' en de voorwaarden uit de bijlage 'Arbeidsvoorwaarden' van toepassing.

Eventuele opmerkingen en/of (wijzigings-)voorstellen op de algemene voorwaarden en op de arbeidsvoorwaarden (heeft overigens niet de voorkeur), dient u zo spoedig mogelijk aan de contactpersoon van het Inkoop Uitvoering Centrum (zie 'communicatie') kenbaar te maken. Op de ingediende reacties zal het Inkoop Uitvoering Centrum reageren en hierbij aangeven of de voorstellen al dan niet worden gehonoreerd.

Eventueel door u gehanteerde (algemene) voorwaarden zijn niet van toepassing.

Als u in uw offerte afwijkt van de (al dan niet aangepaste) ARVODI-2014 en de arbeidsvoorwaarden, zal dit leiden tot het terzijde leggen van uw offerte. Uw offerte komt dan niet voor gunning van de opdracht in aanmerking.

Aan dit offerteverzoek kunt u geen rechten ontlenen op vergoeding van offertekosten of op het verkrijgen van de opdracht.

Door het indienen van een offerte verklaart u te voldoen aan alle in deze offerteaanvraag gestelde eisen en akkoord te gaan met alle hierin gestelde voorwaarden en condities. Indien u hiervan afwijkt, kan dit leiden tot terzijde leggen van uw offerte. Uw offerte komt in dat geval niet voor gunning van de opdracht in aanmerking.

Met vriendelijke groet,



Gea Strootman

## Bijlage 2: Achtergrond wetenschappelijk team

Het wetenschappelijk review team bestaat uit medewerkers van de Universiteit Leiden en van de Universiteit Utrecht.

De betrokken afdelingen en medewerkers van de Universiteit Leiden en van de Universiteit Utrecht hebben een goede (onafhankelijke) reputatie en een groot netwerk, en zijn geen belanghebbende in het aardbevingsdossier. Betrokkenen hebben veel ervaring met het werken in interdisciplinaire projecten, en hebben hier ook veel affiniteit mee. Goed samenwerken, overleggen en communiceren met partijen met een diverse achtergrond maakt hier onderdeel van uit. Het wetenschappelijke team heeft daarnaast ruime ervaring met het werken met gevoelige dossiers (bijvoorbeeld vanuit onderzoeksprogramma's naar CO2 afvang, transport en opslag—Universiteit Leiden: onderzoek naar geluidsoverlast bij Schiphol—Universiteit Utrecht).

Het wetenschappelijke team (Universiteit Leiden, Universiteit Utrecht) heeft uitgebreide methodologische expertise op het gebied van zowel eenmalige als longitudinale surveys (o.a. op het gebied van steekproeftrekking, representativiteit, weging, respons, non-respons, vraagstelling, antwoordmogelijkheden/format en analyses/conclusies). Daarnaast heeft het team uitgebreide sociaalwetenschappelijke expertise op het gebied van publieke acceptatie van energieprojecten en is het bekend met ontwikkelingen in het aardbevingendossier/de regio.

Prof. dr. Jelke Bethlehem is als Emeritus Bijzonder Hoogleraar Survey Methodologie verbonden aan de Universiteit Leiden. Hij heeft meer dan 30 jaar ervaring als senior methodoloog voor het Centraal Bureau voor de Statistiek. Hij geldt wereldwijd als een van de experts op het gebied van het trekken van steekproeven en corrigeren voor fouten die hierin optreden. De laatste jaren heeft hij actief gepubliceerd over de opkomst van web-surveys als een specifieke methode voor survey-onderzoek. Hij schreef diverse handboeken in survey-methodologie, die gelden als standaardwerken, en meer dan 50 artikelen in internationale tijdschriften. Hij is lid van de kerngroep van het Nederlandstalig Platform voor Survey Onderzoek. Daarnaast heeft hij jarenlang advies gegeven aan de EU, ministeries en andere overheidsorganen in binnen- en buitenland.

De afdeling Sociale en Organisationspsychologie van de Universiteit Leiden heeft uitgebreide sociaalwetenschappelijke expertise op het gebied van kwantitatief onderzoek naar factoren die van invloed zijn op publieke acceptatie rondom (on)gewenste energieprojecten (o.a., publieke awareness en kennis, risicoperceptie, tevredenheid met de woonomgeving, vertrouwen in projectontwikkelaars, waargenomen procedurele en distributieve rechtvaardigheid en compensatie van gemeentes die geconfronteerd worden met energieprojecten). Deze expertise wordt internationaal erkend. Zo leidde de Universiteit Leiden van 2004-2014 het sub-programma Publieke Perceptie binnen het nationale multidisciplinaire onderzoeksprogramma naar CO2 afvang, transport en opslag (CATO, totaal 90 miljoen euro, ~3.4 miljoen aan publieke perceptie onderzoek). Hiernaast leidde de Universiteit Leiden van 2009-2010 binnen het multidisciplinaire Europese Fenco Eranet project 'Scrutinizing the impact of CCS communication on the

general and local public' een methodologisch werkpakket waarbinnen middels een grootschalig experimenteel onderzoek de kwaliteit van twee sociaalwetenschappelijke onderzoeksmethoden werd vergeleken.

Dr. Emma ter Mors is als Senior Onderzoeker en Universitair Docent verbonden aan de afdeling Sociale en Organisationspsychologie van de Universiteit Leiden. Ze werkte van 2004-2014 fulltime als onderzoeker en coördinator binnen het CATO project en het Fenco Eranet project en leidde het sub programma Publieke Perceptie binnen het CATO programma van 2012-2014. Ze is sinds 2012 bestuurslid van het International Energy Agency Greenhouse Gas Social Research Network, is sinds 2012 lid van het Expert Review Committee van het tweejaarlijkse internationale Greenhouse Gas Control Technologies congres, en was van 2014-2015 expert lid van de National Council binnen het EU project R&Dialogue (een project gericht op de mogelijke rol van dialoog in energietransitie). Ze publiceert artikelen in internationale toonaangevende tijdschriften en geeft op ad-hoc basis advies aan diverse maatschappelijke partijen (waaronder het Ministerie van Economische Zaken). Ze is vanuit haar CATO onderzoek bekend met het gaswinnings-/aardbevingsdossier en ontwikkelingen binnen in de regio.

De Universiteit Utrecht heeft uitgebreide expertise op het gebied van survey-onderzoek. Het departement Methodologie en Statistiek aan de Faculteit Sociale Wetenschappen is het grootste departement in Europa op het gebied van het opzetten en analyseren van sociaalwetenschappelijk onderzoek.

Dr. Peter Lugtig werkt in de functie van Universitair Hoofddocent Methoden van Onderzoek aan de Universiteit Utrecht. Zijn expertise ligt op het gebied van longitudinaal survey-onderzoek. Hij werkte van 2012-2015 als Senior Onderzoeker aan het Institute of Social and Economic Research aan de University of Essex, en ontving daar een Future Leaders Grant voor zijn onderzoek naar meet- en non-responsfouten in longitudinale surveys. Hij maakt deel uit van het consortium board van het Gender and Generations Programme, waar hij verantwoordelijk is voor de methodologie van het longitudinale survey-onderzoek in 30 Europese landen naar sociologische oorzaken en consequenties van vergrijzing en veranderende gender-rollen. Daarnaast is hij van het methodologische adviescomité van de Engelse Understanding Society survey en is hij lid van de kerngroep van het Nederlands Platform voor Survey Onderzoek. Hij heeft meer dan 30 internationale artikelen gepubliceerd op het gebied van survey-methodologie, en heeft ervaring in het geven van methodologisch advies aan onder meer de Ministeries van V&J, Binnenlandse Zaken, en diverse universiteiten en onderzoeksinstituten.

## Bijlage 3: Longlist en shortlist survey-onderzoeken

### Voorstel voor te reviewen studies—zoals voorgelegd aan de stuurgroep op 19 januari 2017

Emma ter Mors, Peter Lugtig

Samen met NCG hebben we tenminste tien studies geïdentificeerd die geheel of gedeeltelijk gebaseerd zijn op data uit vragenlijstonderzoek. We hebben ons beperkt tot studies die vragen naar hoe mensen in (en soms buiten) Groningen de leefbaarheid ervaren. 'Leefbaarheid' is in de studies soms verschillend gedefinieerd, en meestal worden er in de studies behalve leefbaarheid nog meer concepten gemeten over bijvoorbeeld de achtergrond van de respondenten, woonomgeving, meningen over beleid, gezinssituatie, fysieke en/of mentale gezondheid etc. Wanneer we een studie reviewen, dan beoordelen we de hele studie zolang de data door middel van vragenlijstonderzoek zijn verzameld.

Het betreft hier de volgende studies:

1. Leefbaarometer: is een landelijk model dat inzicht geeft in de leefbaarheid op gedetailleerd niveau; <http://www.leefbaarometer.nl/page/over>
2. WoON: Landelijke enquête die driejaarlijks wordt herhaald en die op regionaal niveau inzicht geeft in diverse indicatoren die op de leefbaarheid betrekking hebben; <https://www.rijksoverheid.nl/onderwerpen/onderzoeken-over-bouwen-wonen-en-leefomgeving/inhoud/lopende-onderzoeken/woononderzoek-nederland-woon>
3. OTB-woningmarktonderzoek aardbevingsgebied: bevat een schat aan informatie over de leefbaarheid; <http://www.otb.tudelft.nl/onderzoek/publicaties/rapporten/>  
[http://www.otb.tudelft.nl/fileadmin/Faculteit/BK/Over\\_de\\_faculteit/Afdelingen/OTB/publicaties/Rapporten/2016\\_A\\_Woningmarkt\\_en\\_leefbaarheidsonderzoek\\_aardbevingsgebied\\_Groningen.pdf](http://www.otb.tudelft.nl/fileadmin/Faculteit/BK/Over_de_faculteit/Afdelingen/OTB/publicaties/Rapporten/2016_A_Woningmarkt_en_leefbaarheidsonderzoek_aardbevingsgebied_Groningen.pdf)
4. Rijksuniversiteit Groningen: <http://www.rug.nl/gmw/psychology/onderzoekgaswinning/> Hoekstra, Elisabeth, Perlaviciute, Goda en Steg, Linda, 2015. Uw mening over gaswinning uit het Groningen-gasveld. Eindrapportage vragenlijstonderzoek 2013-2014. Voorlopige resultaten fase 3 en Hoekstra, Elisabeth, Goda Perlaviciute & Linda Steg, 2014. Uw mening over gaswinning uit het Groningen-gasveld; onderzoeksresultaten fase 1. Rijksuniversiteit Groningen; <http://www.rug.nl/gmw/psychology/onderzoekgaswinning/news-archive/website-document-final.pdf>
5. CMO STAMM, Woon- en Leefbaarheidsmonitor; <http://www.cmostamm.nl/portfolio/onderzoek-naar-woon-en-leefklimaat-in-de-aardbevingsgebieden/>  
<http://www.cmostamm.nl/veiligheid-beleving/>  
<http://www.cmostamm.nl/portfolio/woon-en-leefbaarheidsmonitor-de-marne/>
6. Onderzoek van Sociaal Planbureau Groningen (verschillende onderzoeken en dataverzamelingen, meestal een gelijke onderzoeksopzet volgend):
  - Sociaal Planbureau Groningen, 2013. Factsheet Veiligheidsbeleving. Provincie Groningen; [http://www.sociaalplanbureaugroningen.nl/dynamic/media/1/documents/thema\\_krimp/Factsheet%20-%20](http://www.sociaalplanbureaugroningen.nl/dynamic/media/1/documents/thema_krimp/Factsheet%20-%20)

[Veiligheidsbeleving.pdf](#)

- Sociaal Planbureau Groningen, 2014. Factsheet 'Aardbevingen in Groningen; wat zijn de ervaringen van burgers?' Vragenlijstonderzoek onder leden Groninger Panel in juni 2014; [http://www.sociaalplanbureaugroningen.nl/dynamic/media/1/documents/Groninger\\_Panel/Factsheet\\_Aardbevingen\\_-\\_ervaringen\\_burgers\\_-\\_juni\\_2014.pdf](http://www.sociaalplanbureaugroningen.nl/dynamic/media/1/documents/Groninger_Panel/Factsheet_Aardbevingen_-_ervaringen_burgers_-_juni_2014.pdf).
  - Sociaal Planbureau Groningen, 2014. Aanpakken van de veiligheid; waar ligt de prioriteit; <http://www.sociaalplanbureaugroningen.nl/veiligheid/prioriteit-veiligheid/>
  - Sociaal Planbureau Groningen, 2014. Factsheet Invloed van Groningse aardbevingen op woongenot; <http://www.sociaalplanbureaugroningen.nl/page/210/aardbevingen/woongenot.html>
  - Sociaal Planbureau Groningen, 2015a. Factsheet Belevingen beeldvorming van het Groninger landschap; <http://www.sociaalplanbureaugroningen.nl/factsheet-beeldvorming-landschap/>
  - Sociaal Planbureau Groningen, 2015b. Factsheet: Voelen Groningers zich verbonden met de provincie? <http://www.sociaalplanbureaugroningen.nl/trots-op-groningen/binding-met-groningen/>
  - Sociaal Planbureau Groningen, 2015. Cijfers veilig woon- en leefomgeving, thema huiselijk geweld; <http://www.sociaalplanbureaugroningen.nl/swing--veilige-leef-en-woonomgeving/>.
  - Sociaal Planbureau Groningen, 2016. Leefbaarheid en bevolkingskrimp: Groningers waarderen leefbaarheid positief, al zien ze wel achteruitgang <http://www.sociaalplanbureaugroningen.nl/ervaren-leefbaarheid/leefbaarheid-positief/>
7. Veiligheidsmonitor: betreft landelijke monitor die inzicht geeft in de waardering die mensen hebben over de leefbaarheid van hun woonbuurt; <http://www.veiligheidsmonitor.nl/>
  8. RIGO-onderzoek uit 2011 naar Leefbaarheid in krimpgebieden (een verkenning van de relatie tussen bevolkingskrimp en leefbaarheid); [http://www.rigo.nl/NL/Publicaties/Publicatie/\\_p/itemID/2587/Leefbaarheid-in-krimpgebieden.aspx](http://www.rigo.nl/NL/Publicaties/Publicatie/_p/itemID/2587/Leefbaarheid-in-krimpgebieden.aspx) en [http://kkn.vanmeernaarbeter.nl/sites/kkn.vanmeernaarbeter.nl/files/Leefbaarheid%20in%20krimpgebieden%20eindversie%20definitief\\_0.pdf](http://kkn.vanmeernaarbeter.nl/sites/kkn.vanmeernaarbeter.nl/files/Leefbaarheid%20in%20krimpgebieden%20eindversie%20definitief_0.pdf).
  9. Royal Haskoning/DHV, Maatschappelijke effect inventarisatie, 2016; [http://www.nam.nl/nam-en-de-samenleving/voorkomen-beperken-overlast/onderzoek-negatieve-gevolgen-gaswinning/\\_jcr\\_content/par/textimage.ce7470e6b33f1b9e0b5667976556e8ddf68073b2a24a9402f8/rapportage-2015-maatschappelijke-effecten-inventarisatie-11022016.pdf](http://www.nam.nl/nam-en-de-samenleving/voorkomen-beperken-overlast/onderzoek-negatieve-gevolgen-gaswinning/_jcr_content/par/textimage.ce7470e6b33f1b9e0b5667976556e8ddf68073b2a24a9402f8/rapportage-2015-maatschappelijke-effecten-inventarisatie-11022016.pdf)
  10. APE Public Economics, 2015. Onderzoek naar de tevredenheid met de schadeafhandeling door de NAM, rapportnr. 1281. Onderzoek in opdracht van de Dialoogtafel Groningen. Den Haag, augustus 2015; <http://www.ape.nl/nl/projecten/default.asp?nID=226>.

Na het bekijken van de documentatie van deze onderzoeken (zie URL's) hebben we als onderzoeksteam een voorkeur voor het reviewen van de volgende vier studies. We hebben voor deze studies gekozen, omdat ze verschillende opdrachtgevers hebben en ook verschillen in de opzet van het onderzoek:

- Leefbaarometer/ WoON (studie 1 of 2).  
Goed om een landelijk onderzoek mee te nemen. Het is wel nodig dat er een regionaal rapport is (denken wel dat dit er is) om onze review uit te kunnen

- voeren. Opdrachtgever: Centraal Bureau Statistiek, ministerie van Wonen.
- RUG-studie van Hoekstra, Perlaviciute, en Steg, Linda (studie 4).
- Onderzoek van Sociaal Planbureau Groningen (SPG) (studie 6).  
SPG heeft wel meerdere survey-onderzoeken gedaan, en we zullen er daar één onderzoek specifiek van kiezen dat representatief is voor de andere onderzoeken. Opdrachtgever: SPG.
- CMO STAMM (studie 5).  
Deze onderzoeken zijn soms onderdeel van het OTB-onderzoek (gebruiken zelfde data), en worden soms samen uitgevoerd met het Sociaal Planbureau Groningen. Het OTB-rapport (grotere onderzoek) heeft twee deelonderzoeken die geschikt zijn voor de review. Hier zouden we ons op richten dan. Betreft surveys onder inwoners. Opdrachtgever: gemeenten/sociaal planbureau.

Naast deze vier onderzoeken hebben we tenminste nog ruimte om één extra onderzoek te reviewen.

Tijdens de eerste vergadering met de stuurgroep op 19 januari is de longlist, alsmede de shortlist met de voorkeur van NCG en het wetenschappelijk team voorgelegd aan de stuurgroep. Op basis van een discussie met de stuurgroep is er uiteindelijk gekozen voor het reviewen van de volgende vijf studies:

- 'Gronings Perspectief', uitgevoerd door de Rijksuniversiteit Groningen in samenwerking met de GGD en Onderzoeksbureau Onderzoek en Statistiek Groningen.
- 'Wonen en aardbevingen in Groningen', uitgevoerd door OTB Delft in samenwerking met CMO STAMM.
- De Leefbaarometer 2.0 in combinatie met WoOn, uitgevoerd door Atlas voor Gemeenten in samenwerking met RIGO.
- Het onderzoek 'Leefbaarheid en bevolkingskrimp in Groningen', uitgevoerd met behulp van het Groninger Panel door het Sociaal Planbureau Groningen.
- 'Onderzoek naar de tevredenheid met de schadeafhandeling door de NAM', uitgevoerd door onderzoeksbureau APE Public Economics.

# Bijlage 4: Een technisch overzicht van het trekken van steekproeven, en correctiemethodes voor non-respons door middel van weging

In deze bijlage wordt op een meer technische wijze uitgewerkt op welke wijze door middel van het trekken van steekproeven uitspraken kunnen worden gedaan over een populatie. Wij vinden het van belang deze bijlage op te nemen, omdat we in onze aanbevelingen een lans breken voor het trekken van een aselecte steekproef, met een complex (gestratificeerd) design. Het schatten van effecten in dit soort designs met non-respons is ingewikkeld. De bijlage geeft een eerste inzicht in hoe dit soort steekproeven werken, en bevat referenties naar aanvullende literatuur waarin op een uitgebreidere manier staat beschreven hoe om te gaan met complexe steekproefdesigns, non-respons en de correctie daarvoor in analyses.

## Doelpopulatie

In statistische termen wordt de doelpopulatie meestal aangegeven met  $U$ . Dat is de  $U$  van universum. De omvang van de doelpopulatie wordt genoteerd met  $N$ . De leden van de doelpopulatie kunnen dus worden genummerd van 1 t/m  $N$ .

In de survey worden variabelen gemeten. Er zijn doelvariabelen en hulpvariabelen. De doelvariabelen zijn de belangrijke variabelen van het onderzoek. Een willekeurige doelvariabele wordt aangeduid met  $Y$ . Voor elke persoon in de doelpopulatie heeft de doelvariabele een waarde. Al die  $N$  waarden worden aangegeven met  $Y_1, Y_2, \dots, Y_N$ .

Hulpvariabelen worden gebruikt om de uitkomsten van de survey nader te differentiëren. Als geslacht bijvoorbeeld een hulpvariabele is, dan kan worden vastgesteld of de uitkomsten voor vrouwen anders zijn dan die voor mannen. Hulpvariabelen kunnen ook worden gebruikt om de non-respons te analyseren en om de vertekening ten gevolge van non-respons te reduceren. Een willekeurige hulpvariabele wordt aangeduid met  $X$ . Voor elke persoon in de doelpopulatie heeft de hulpvariabele een waarde. Al die  $N$  waarden worden aangegeven met  $X_1, X_2, \dots, X_N$ .

Het schatten van allerlei karakteristieken van de doelpopulatie zal meestal het doel van de survey zijn. Bij een continue variabele  $Y$  (bijvoorbeeld omvang van de schade van een persoon) zou dat bijvoorbeeld het gemiddelde in de populatie kunnen zijn. Dit wordt genoteerd met:

$$\bar{Y} = \frac{Y_1 + Y_2 + \dots + Y_N}{N} = \frac{1}{N} \sum_{k=1}^N Y_k \quad (1)$$

Een doelvariabele kan ook de aan- of afwezigheid van een bepaald kenmerk meten (bijvoorbeeld het al of niet hebben van schade). De doelvariabele krijgt dan de waarde 1 als de desbetreffende persoon het kenmerk wel heeft ( $Y_k = 1$ ) en de waarde 0 als dat niet zo is ( $Y_k = 0$ ). Het percentage personen in de doelpopulatie met dat kenmerk kan

dan worden genoteerd met:

$$P = 100 \times \bar{Y} = 100 \times \frac{Y_1 + Y_2 + \dots + Y_N}{N} = \frac{100}{N} \sum_{k=1}^N Y_k \quad (2)$$

### Steekproef trekken

Laat  $n$  de omvang van de steekproef aanduiden. Merk op dat grootheden die betrekking hebben op de populatie worden aangeduid met hoofdletters, terwijl steekproefgrootheden worden aangeduid met kleine letters. De steekproef kan nu worden aangeduid met een reeks van  $N$  indicatoren  $a_1, a_2, \dots, a_N$ , waarbij  $a_k = 1$  als persoon  $k$  in de steekproef zit, en  $a_k = 0$  als dit niet het geval is. Daarbij loopt  $k$  van 1 tot  $N$ . De omvang  $n$  van de steekproef kan worden geschreven als:

$$n = a_1 + a_2 + \dots + a_N = \sum_{k=1}^N a_k \quad (3)$$

Het loten van de steekproef heeft twee belangrijke voordelen. In de eerste plaats kunnen dan altijd valide schatters (d.w.z. zuivere schatters) worden bepaald van allerlei kenmerken in de populatie. In de tweede plaats is het mogelijk om te kwantificeren hoe dicht de schattingen bij de werkelijke (onbekende) populatiewaarden in de buurt liggen. Dit gebeurt meestal in de vorm van betrouwbaarheidsintervallen of onzekerheidsmarges.

Al in 1952 toonden Horvitz en Thompson hoe een zuivere schatter kan worden geconstrueerd. De selectiekansen van de  $N$  personen in de doelpopulatie worden aangegeven met  $\pi_1, \pi_2, \dots, \pi_N$ . Die selectiekansen kunnen eventueel ongelijk zijn. Ongelijke selectiekansen komen voor in de gereviewde Groningse surveys. Soms hebben bewoners van dunbevolkte gebieden een grotere selectiekans dan bewoners in dichtbevolkte gebieden. De steekproef wordt aangegeven met de reeks steekproefindicatoren  $a_1, a_2, \dots, a_N$ . Dan volgt uit de theorie van Horvitz en Thompson (1952) dat

$$\bar{y}_{HT} = \frac{1}{N} \times \left( a_1 \frac{Y_1}{\pi_1} + a_2 \frac{Y_2}{\pi_2} + \dots + a_N \frac{Y_N}{\pi_N} \right) = \frac{1}{N} \sum_{k=1}^N a_k \frac{Y_k}{\pi_k} \quad (4)$$

een zuivere schatter is voor het populatiegemiddelde van een continue variabele  $Y$ . Merk op dat wordt gewogen met de selectiekansen om zo te kunnen corrigeren voor de oververtegenwoordiging van personen met grote selectiekansen. Verder uitwerken van de theorie wijst ook uit dat bij gelijke selectiekansen ( $\pi_k = n/N$ , voor alle  $k$ ) het steekproefgemiddelde een zuivere schatter is voor het populatiegemiddelde en dat het steekproefpercentage een zuivere schatter is voor het populatiepercentage.

Als de steekproef is geloot uit de doelpopulatie, kan ook de nauwkeurigheid van de schattingen worden bepaald. De variantie van de Horvitz-Thompsonschatter is gelijk aan:

$$V(\bar{y}_{HT}) = \frac{1}{2N^2} \sum_{k=1}^N \sum_{l=1}^N (\pi_k \pi_k - \pi_{kl}) \left( \frac{Y_k}{\pi_k} - \frac{Y_l}{\pi_l} \right)^2 \quad (5)$$

Hierin is  $\pi_{kl}$  de kans dat de twee elementen  $k$  en  $l$  samen in de steekproef worden geselecteerd. In het speciale geval van gelijke trekkingskansen en het schatten van het populatiegemiddelde van een continue variabele, reduceert deze uitdrukking tot:

$$V(\bar{y}) = \frac{1-n/N}{n} S^2 = \frac{1-n/N}{n} \frac{1}{N} \sum (Y - \bar{Y})^2. \quad (6)$$

Hieraan is bijvoorbeeld duidelijk te zien dat de nauwkeurigheid van de schattingen toeneemt met de omvang van de steekproef. In geval van gelijke trekkingskansen en het schatten van een populatiepercentage  $P$  is de variantie van het steekproefpercentage  $p$  gelijk aan:

$$V(p) = \frac{1-n/N}{n} S^2 = \frac{1-n/N}{n} P(100-P). \quad (7)$$

De variantie kan worden gebruikt om het 95%-betrouwbaarheidsinterval te bepalen. Eerst wordt de standaardfout van de schatter uitgerekend. Die is gelijk aan de wortel uit de variantie van de schatter. Voor het steekproefpercentage is dit:

$$S(p) = \sqrt{V(p)}. \quad (8)$$

De onzekerheidsmarge van het 95%-betrouwbaarheidsinterval wordt nu verkregen door de standaardfout met 1,96 te vermenigvuldigen:

$$M(p) = 1.96 \times S(p). \quad (9)$$

De interpretatie van de onzekerheidsmarge is dat met een grote mate van waarschijnlijkheid (95%) het werkelijk percentage in de populatie niet meer dan  $M(p)$  afwijkt van het steekproefpercentage.

Het 95%-betrouwbaarheidsinterval is nu gelijk aan:

$$CI(p) = (p - M(p) ; p + M(p)). \quad (10)$$

De interpretatie van dit interval is dat met een waarschijnlijkheid van 95% dit interval de werkelijke populatiewaarde bevat.

De hierboven beschreven theorie is dus alleen toepasbaar als de steekproef door loting is verkregen.

### *Meer geavanceerde steekproefdesigns*

Wanneer een enkelvoudig aselecte steekproef wordt getrokken, dan heeft iedereen in de doelpopulatie (mits er een goede lijst wordt gebruikt) dezelfde kans om in de steekproef terecht te komen. Hoewel dit voor veel typen onderzoek een wenselijk design is, zijn er goede redenen te kiezen voor een complexer design van de steekproef. Twee soorten steekproeven zijn relevant als het gaat om onderzoek naar de leefbaarheid in Groningen.

- **De gestratificeerde aselechte steekproef:**  
Om ervoor te zorgen dat er van een bepaalde groep personen voldoende mensen in de steekproef belanden, kan ervoor worden gekozen deze mensen een grotere kans te geven in de steekproef te komen. Als de doelpopulatie bestaat uit volwassenen die wonen in de provincie Groningen, dan kan er bijvoorbeeld worden gestratificeerd op urbanisatiegraad (d.w.z. plattelandsgebieden krijgen een hogere kans), op gemeenten (d.w.z. bewoners van kleine gemeenten krijgen een hogere kans), of nabijheid bij het centrum van het aardbevingsgebied (d.w.z. bewoners uit Loppersum krijgen een hogere kans dan bewoners uit bijvoorbeeld ter Apel). Er kan theoretisch ook worden gestratificeerd op andere kenmerken, zoals leeftijd, geslacht, of type huis. Of en op welke manier stratificatie zinvol is hangt af van twee mogelijke doelen. Ten eerste kan stratificatie ervoor zorgen dat relatief kleine groepen mensen in voldoende mate worden betrokken bij het onderzoek (bijvoorbeeld boerderijbewoners). Een tweede doel kan zijn de steekproef efficiënter te maken. Wanneer wordt gestratificeerd op variabelen die sterk samenhangen met de doelvariabelen (leefbaarheid), dan kunnen verschillen in ervaren leefbaarheid worden aangetoond met een relatief kleinere steekproef (zie ook volgende sectie).
- **De clustersteekproef:**  
In deze steekproef wordt vanwege financiële overwegingen eerst een steekproef getrokken van geografische gebieden. Zo is het bij een huis-aan-huis survey tijdrovend (en daardoor duur) om kriskras door Groningen te reizen en een vragenlijst af te nemen. Het is goedkoper om eerst een aantal gebieden (bijvoorbeeld gemeenten of kernen) aselechte te kiezen, en vervolgens alleen in die gebieden onderzoek te doen, meestal door in die kernen weer een steekproef te trekken. De door de clustersteekproef bespaarde kosten kunnen vervolgens worden gebruikt om de steekproef te vergroten, en zo meer precisie of zekerheid over de schattingen te krijgen. Ook bij clustersteekproeven zijn de selectiekansen vaak niet gelijk. In de analyse moet daarmee rekening worden gehouden, en worden gecorrigeerd voor deze ongelijke trekkingskansen door het maken van weegfactoren (designgewichten), die de scheve selectiekansen weer gelijk trekken.

### *Hoe groot moet de steekproef zijn?*

Als de steekproef door loting is verkregen, dan kan de nauwkeurigheid van de schattingen worden berekend. De nauwkeurigheid wordt meestal weergegeven in de vorm van de onzekerheidsmarge. Die onzekerheidsmarge geeft aan hoeveel de schatting maximaal kan afwijken van de werkelijke waarde in de doelpopulatie als gevolg van de onzekerheid in de resultaten door de rol van toeval in het trekken van de steekproef.

Voor grote doelpopulaties kan bijvoorbeeld de onzekerheidsmarge voor het steekproefpercentage in een aselechte steekproef met gelijke kansen benaderd worden met:

$$M(p) = 1,96 \times \sqrt{\frac{P(100-P)}{n}}. \quad (11)$$

Uit deze formule kan worden afgeleid dat de onzekerheidsmarge kleiner wordt (en de schattingen dus nauwkeuriger) als de omvang van de steekproef groter wordt. Ook

is te zien dat de nauwkeurigheid niet afhangt van de omvang van de doelpopulatie. Voor een grotere doelpopulatie is dus niet een grotere steekproef nodig om dezelfde nauwkeurigheid te verkrijgen. Tabel 3.1 bevat voor verschillende steekproefgroottes en een reeks van steekproefpercentages de bijbehorende onzekerheidsmarge.

Tabel 3.1. Onzekerheidsmarges

Gevonden percentage resultaat in de steekproef als	Omvang van de steekproef (n)					
	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

Een voorbeeld: Stel dat uit een peiling onder een steekproef van 500 personen blijkt dat 40% van de respondenten een bepaalde eigenschap heeft (bijvoorbeeld schade aan het huis). Bij die 500 en die 40% hoort een marge van 4,3%. Het percentage respondenten met schade aan het huis zal met grote waarschijnlijkheid dus liggen tussen  $(40-4,3 =) 35,7%$  en  $(40+4,3 =) 44,3%$ .

Een tweede voorbeeld: Stel dat in een peiling onder 1000 personen wordt gevraagd of mensen overwegen te verhuizen. Uit de peiling blijkt dat 20% dat overweegt. Een maand later wordt een nieuwe peiling uitgevoerd, ook onder 1000 personen. Nu blijkt 22% overweegt te verhuizen. Kun je nu zeggen dat er meer mensen zijn die overwegen te verhuizen? Nee, want beide percentages hebben een onzekerheidsmarge van 2,5%. Die marge is groter dan het verschil tussen de percentages mogelijke verhuizers (nl.  $22 - 20 = 2%$ ). Het verschil kan dus net zo goed veroorzaakt door de 'toeval' in de steekproef. Er is geen sprake van een 'echt' verschil.

### Non-respons

Non-respons wordt vaak gemodelleerd door aan iedere persoon in de populatie een onbekende kans op respons toe te kennen. De responskansen worden aangegeven met  $\rho_1, \rho_2, \dots, \rho_N$ . Als een persoon  $k$  in de doelpopulatie wordt gevraagd om mee te doen in de survey, dan zal dat met kans  $\rho_k$  gebeuren, en met kans  $1-\rho_k$  zal er non-respons optreden.

De vertekening als gevolg van non-respons kan ook statistisch worden uitgedrukt. De vertekening (ofwel *bias*) is dan gelijk aan:

$$Bias = \frac{R_{Y,\rho} \times S_Y \times S_\rho}{\bar{\rho}} \quad (12)$$

Hier is  $R_{y,p}$  de correlatiecoëfficiënt. Dat is een maat voor de samenhang tussen de uitkomstvariabele en de responskansen. De correlatiecoëfficiënt is alleen nul als er geen samenhang is. Dus een samenhang draagt bij aan de vertekening.

Verder is  $S_p$  de wortel uit de variantie van de responskansen. Naarmate de responskansen meer variëren (er zijn grote kansen en kleine kansen), zal de vertekening groter zijn. Als alle responskansen gelijk zijn, is er geen vertekening.

De grootte is het gemiddelde van alle responskansen. Deze grootte kan worden geschat met de fractie respons in de survey. Een hoge fractie respons betekent een grote gemiddelde responskans en dus een kleine vertekening. Bij een lage responsfractie is de vertekening groot.

### *Correctie voor non-respons*

Een eenvoudig voorbeeld: de respons van een survey onder de algemene bevolking blijkt 60% uit mannen bestaan te bestaan en dus voor 40% uit vrouwen. Dat is niet goed, want de Nederlandse bevolking bestaat voor 49,5% uit mannen en voor 50,5% uit vrouwen. Om hiervoor te corrigeren, krijgt elke responderende man een gewicht van:

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

Dit is het populatiepercentage gedeeld door het percentage in de respons. Dus elke man telt nog maar mee voor 0,825 man in plaats van voor 1 man. Het gewicht is kleiner dan 1 omdat er te veel mannen waren in de respons. Elk vrouw krijgt een gewicht van:

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

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

Het toekennen van gewichten aan mannen en vrouwen leidt ertoe dat de resultaten van de survey 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 hulpvariabelen door de resultaten te corrigeren met een weegfactor. Noodgedwongen kunnen dat alleen kenmerken zijn die in de survey zijn gemeten, en waarvoor de verdeling in de populatie bekend is. Voorbeelden van veel gebruikte kenmerken zijn geslacht, leeftijd, burgerlijke staat, etniciteit, regio en opleidingsniveau.

De hoop is dat als de resultaten van de survey representatief zijn gemaakt met betrekking tot veel hulpvariabelen, deze misschien ook wel representatief worden met betrekking tot de uitkomsten van het onderzoek (bijvoorbeeld leefbaarheid). Daarmee zou dan de vertekening door non-respons 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 relatie tussen de doelvariabelen en de weegvariabelen (de hulpvariabelen die in de weging worden gebruikt).

- Er is een sterke relatie tussen het antwoordgedrag in het onderzoek (wie doen er wel en niet mee) en de weegvariabelen.

Soms heeft wegen veel effect op de uitkomsten van een onderzoek, 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 zowel invloed op de tijdbesteding, en ze verklaren waarom mensen vaker meedoen aan de survey. Maar het wegen met leeftijd en geslacht in een onderzoek naar meningen is vaak zinloos, omdat die kenmerken niet sterk samenhangen met meningen.

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 een goede oplossing is om die vertekening op te lossen. Er bestaan verschillende technieken om te wegen. Welke methode wordt gebruikt hangt af van de manier waarop informatie over de populatie (in de vorm van hulpvariabelen) beschikbaar is.

De eenvoudigste en meest gebruikte weegtechniek is *post-stratificatie*. Hierbij wordt de doelpopulatie in groepen (strata) verdeeld op basis van de hulpvariabelen. Alle respondenten in een stratum krijgen hetzelfde gewicht. Dit gewicht is gelijk aan het percentage respondenten in het stratum gedeeld door het percentage personen in de doelpopulatie dat in het stratum zit.

Stel dat de hulpvariabelen de doelpopulatie  $U$  verdelen in  $L$  strata  $U_1, U_2, \dots, U_L$  met omvang  $N_1, N_2, \dots, N_L$ , waarbij  $N = N_1 + N_2 + \dots + N_L$ . Veronderstel dat een aselechte steekproef met gelijke kansen is getrokken. De schatting voor het populatiegemiddelde van de doelvariabele  $Y$  is, na post-stratificatie, gelijk aan:

$$\bar{y}_{PS} = \frac{N_1}{N} \bar{y}^{(1)} + \frac{N_2}{N} \bar{y}^{(2)} + \dots + \frac{N_L}{N} \bar{y}^{(L)} = \frac{1}{N} \sum_{h=1}^L N_h \bar{y}^{(h)}. \quad (13)$$

Hierin zijn de responsgemiddelden van de doelvariabele in de  $L$  strata. Post-stratificatie werkt vooral als de strata homogeen zijn. Dat wil zeggen dat de respondenten in een stratum veel op elkaar lijken (met betrekking tot responsgedrag en de doelvariabelen).

Wegen corrigeert beter naarmate meer effectieve weegvariabelen worden gebruikt. Daar is echter een grens aan. Als het aantal strata erg groot wordt, kunnen er lege strata ontstaan, strata zonder waarnemingen. Dan kunnen geen gewichten meer worden berekenen, en kan dus ook niet worden gecorrigeerd voor de vertekening ten gevolge van non-respons. Er zijn alternatieve weegtechnieken beschikbaar die kunnen worden toegepast daar waar post-stratificatie niet meer mogelijk is.

De eerste alternatieve techniek is *lineair wegen*. Deze techniek wordt ook wel *generalized regression estimation* genoemd. Uitgangspunt van lineair wegen is een lineair regressiemodel dat de waarden van een doelvariabele probeert te schatten op basis van een reeks hulpvariabelen. Dit leidt tot de volgende regressieschatter:

$$\bar{y}_{REG} = \bar{y} + (\bar{X} - \bar{x})' b \quad (14)$$

Hierin is het steekproefgemiddelde van de doelvariabele, is de vector van populatiegemiddelden van de hulpvariabele, is de vector of responsgemiddelden van

de hulpvariabelen, en is  $b$  een vector van regressiecoëfficiënten. De regressieschatting kan worden gezien als een correctie van het simpele steekproefgemiddelde, waarbij gecorrigeerd wordt voor een te groot of te klein steekproefgemiddelde van de hulpvariabelen.

De vertekening van de regressieschatting is kleiner naarmate het regressiemodel beter past. Er kan worden aangetoond dat post-stratificatie een speciaal geval is van de regressieschatting.

Een tweede alternatieve techniek voor wegen is *multiplicatief wegen*. Deze techniek wordt ook wel *raking*, *raking ratio estimation*, *iterative proportional fitting* of *RIM weighting* genoemd. De techniek komt neer op het tegelijk uitvoeren van een aantal post-stratificaties. Dit gebeurt in een iteratief proces, waarbij de post-stratificaties achter elkaar worden uitgevoerd. Een post-stratificatie past de uitkomst van de voorgaande post-stratificaties aan. Toch is het zo dat na het een aantal herhalingen van de reeks post-stratificaties het proces convergeert.

De uitkomsten van lineair wegen en multiplicatief wegen lijken asymptotisch op elkaar. Lineair wegen heeft het voordeel dat de varianties van de gewogen schatters simpel kunnen worden berekend. Bij multiplicatief wegen kan dit niet. Lineair wegen heeft het nadeel dat gewichten negatief kunnen zijn. Bij multiplicatief wegen zijn de gewichten altijd positief.

Er is ook nog een heel andere aanpak en die is gebaseerd op het werken met responskansen. In het geval van non-respons is de aangepaste Horvitz-Thompson-schatting

$$\bar{y}_{HT,R} = \frac{1}{N} \times \left( a_1 R_1 \frac{Y_1}{\pi_1 \rho_1} + a_2 R_2 \frac{Y_2}{\pi_2 \rho_2} + \dots + a_N R_N \frac{Y_N}{\pi_N \rho_N} \right) = \frac{1}{N} \sum_{k=1}^N a_k R_k \frac{Y_k}{\rho_k \pi_k} \quad (15)$$

een zuivere schatter. Hierin zijn  $R_1, R_2, \dots, R_N$  indicatoren voor het wel ( $R_k = 1$ ) of niet ( $R_k = 0$ ) responderen van personen. Deze schatter kan niet worden gebruikt omdat de responskansen onbekend zijn. De oplossing is dan de responskansen te schatten. Meestal wordt hiervoor een logistisch model gebruikt.

$$\text{logit}(\rho_k(X)) = \log\left(\frac{\rho_k(X)}{1 - \rho_k(X)}\right) = X_k' \beta. \quad (16)$$

Hierin is  $X_k$  een vector van waarden van de hulpvariabelen, en  $\beta$  is een vector van regressiecoëfficiënten. Gecorrigeerde schattingen kunnen nu worden verkregen door de responskansen  $\rho_k$  in formule (14) te vervangen door geschatte kansen:

$$\hat{y}_{HT,R} = \frac{1}{N} \sum_{k=1}^N a_k R_k \frac{Y_k}{\pi_k \hat{\rho}_k(X)}. \quad (17)$$

Deze aangepaste Horvitz-Thompson-schatting werkt beter naarmate het logistische model (15) beter in staat is om de responskansen te schatten.

Als geschatte responskansen beschikbaar zijn, kan ook nog worden besloten om ze op een andere manier te gebruiken. Het is namelijk ook mogelijk om een post-stratificatie

uit te voeren waarbij de geschatte responskansen als stratificatie-variabele worden gebruikt. Door de doelpopulatie te verdelen in strata die homogeen zijn met betrekking tot de responskansen, kan de vertekening in de uitkomsten worden weggewerkt. Ook hier geldt weer dat dit beter werkt naarmate de responskansen beter kunnen worden geschat.

Uitgebreidere informatie over verschillende weegtechnieken kan bijvoorbeeld worden gevonden in Bethlehem, Cobben, en Schouten (2011).

Vooraf wanneer er sprake is van een substantiële hoeveelheid non-respons, moet er rekening worden gehouden met ernstige vertekeningen in de uitkomsten. Daarom is het van groot belang de uitkomsten te wegen. Daarvoor zijn voldoende effectieve weegvariabelen nodig. Helaas zijn die ingrediënten lang niet altijd aanwezig. Meestal kan wel worden gewogen naar variabelen als leeftijd, geslacht, regio en opleiding, maar die variabelen zijn lang niet altijd voldoende om een vertekening weg te werken.

# Bijlage 5: Overzicht van onderzoeken naar leefbaarheid in aardbevingsgebied Groningen

## Gereviewde onderzoeken:

- Boelhouwer, P., Boumeester, H., Groetelaers, D., Hoekstra, J., VanderHeijden, H., Jansen, S., ... Ringersma, R. (2016, januari). *Woningmarkt- en leefbaarheidsonderzoek aardbevingsgebied Groningen*. Onderzoek voor de gebouwde omgeving (OTB) & CMO/STAMM.
- Cremers, M. & Engelen, F. (2014, 18 juni). *Woononderzoek Nederland (WoON) 2015*. Centraal Bureau voor de Statistiek.
- Cremers, M., & Engelen, F. (2016, 26 januari). *Woononderzoek Nederland (WoON) 2015*. Centraal Bureau voor de Statistiek.
- Faber, D., Van der Veen, S., De Jong, P. (2015, augustus). *Onderzoek naar de tevredenheid met de schadeafhandeling door de NAM*. APE.
- Hoekstra, J. (2016, januari). *Wonen en aardbevingen in Groningen: Een onderzoek in negen gemeenten*. Onderzoek voor de gebouwde omgeving (OTB).
- Hoekstra, J. (2016, januari). *Wonen en aardbevingen in Groningen: Een onderzoek in negen gemeenten, Bijlagenboek*. Onderzoek voor de gebouwde omgeving (OTB).
- Janssen, S. (2016, april 1). *WoON2015 Onderzoek documentatie: Dataprocesing*. Centraal Bureau voor de Statistiek.
- Leidelmeijer, K., Marlet, G., Ponds, R., Schulenberg, R. & Woerkens, C. (2014). *Leefbaarheid in beeld: Analyse van de leefbaarheidsontwikkeling 2012-2014 volgens de Leefbaarometer 2.0*. Atlas voor gemeenten.
- Leidelmeijer, K., Marlet, G., Ponds, R., Schulenberg, R. & Woerkens, C. (2014). *Leefbaarometer 2.0: Instrumentontwikkeling*. Atlas voor gemeenten.
- Postmes, T., Stroebe, K., Richardson, J., LeKander, B., & Oldersma, F (2016, april). *Tussenrapport Gronings Perspectief #1*. Rijksuniversiteit Groningen.
- Postmes, T., Stroebe, K., Richardson, J., LeKander, B., & Oldersma, F (2016, oktober). *Tussenrapport Gronings Perspectief #2*. Rijksuniversiteit Groningen.
- Postmes, T., Stroebe, K., Richardson, J., LeKander, B., & Oldersma, F (2017, 7 februari). *Bijlagen bij "Veiligheidsbeleving, gezondheid en toekomstperspectief van Groningers"*. Rijksuniversiteit Groningen.
- Postmes, T., Stroebe, K., Richardson, J., LeKander, B., & Oldersma, F. (2016, 15 juli). *Veiligheidsbeleving, gezondheid en toekomstperspectief van Groningers: Wetenschappelijk rapport #1*. Rijksuniversiteit Groningen.
- Postmes, T., Stroebe, K., Richardson, J., LeKander, B., & Oldersma, F. (2017, 7 februari). *Veiligheidsbeleving, gezondheid en toekomstperspectief van Groningers: Wetenschappelijk rapport #2*. Rijksuniversiteit Groningen.

Postmes, T., Stroebe, K., Richardson, J., LeKander, B., & Oldersma, F. (2016, 29 april). *Verantwoording eerste meting panelonderzoek Gronings Perspectief*. Rijksuniversiteit Groningen.

Sociaal Planbureau Groningen, 2016. Leefbaarheid en bevolkingskrimp: Groningers waarderen leefbaarheid positief, al zien ze wel achteruitgang <http://www.sociaalplanbureaugroningen.nl/ervaren-leefbaarheid/leefbaarheid-positief/>

Vondenhoff, E. (2015). *Weging WOON 2015*. Centraal Bureau voor de Statistiek.

### **Indicatief overzicht van niet gereviewde kwantitatieve en kwalitatieve onderzoeken:**

Anoniem (2015, 18 maart). Solidair met gedupeerden gaswinning. *Nederlands Dagblad*. Verkregen op: <https://www.nd.nl/nieuws/gratis-nieuws/solidair-met-gedupeerden-gaswinning.462038.lynkx>

Anoniem (2017, 29 maart). Kinderombudsvrouw be vraagt kinderen over gaswinning. *NOS*. Verkregen op: <http://nos.nl/artikel/2165512-kinderombudsvrouw-bevraagt-kinderen-over-gaswinning.html>

Briene, M., & Meurs, E. (2017). *Praktische uitwerking schadeprotocol Centrum Veilig Wonen*. Nationaal Coördinator Groningen.

Commissie Duurzame Toekomst Noord-Oost Groningen (2013). *Vertrouwen in een duurzame toekomst: Een stevig perspectief voor Noord-Oost Groningen*.

Drenth, E. (2013). Helft stemmers poll Groningen dichtbij voor stoppen gaswinning. *Dichtbij*. Verkregen op: <http://www.dichtbij.nl/groningen/regionaal-nieuws/artikel/2634557/helft-stemmers-poll-groningen-dichtbij-voor-stoppen-gaswinning.aspx>

Groninger Bodem Beweging (2013, 27 maart). Poll gesloten. Verkregen op: <http://www.groninger-bodem-beweging.nl/gbb?catid=0&id=367>

Hoekstra, E., Perlaviciute, G., & Steg, L. (2015). *Uw mening over gaswinning uit het Groningen-gasveld: Onderzoeksresultaten fase 1, 2, en 3*. Rijksuniversiteit Groningen.

Meinders, K. (2016, april 18). NAM wil weer meer gas gaan oppompen in Groningen. *BNR*. Verkregen op: <https://www.bnr.nl/nieuws/10016410/nam-wil-weer-meer-gas-gaan-oppompen-in-groningen>

Onderzoeksraad voor Veiligheid (2015, februari). *Aardbevingsrisico's in Groningen: Onderzoek naar de rol van veiligheid van burgers in de besluitvorming over de gaswinning (1959-2014)*.

Provincie Groningen (2017). *Alles op een rij: Een chronologisch overzicht van het aardbevingsdossier*. Verkregen op: <https://www.provinciegroningen.nl/actueel/dossiers/aardbevingen/alles-op-een-rij/>

Royal HaskoningDHV (2017, 1 maart). *Finaal concept-rapport: Bijlagen - Maatschappelijke effecten inventarisatie van aardbevingen in Noordoost-Groningen 2016*. Verkregen op: <http://www.nam.nl/nam-en-de-samenleving/voorkomen-beperken-overlast/onderzoek-negatieve->

gevolgen-gaswinning/\_jcr\_content/par/textimage.stream/1490092619035/d27c67d34f2c4565b4e3a1aaab0ab5795e687cc6913e959c304bc49/bijlagen-maatschappelijke-effecten-inventarisatie-2016.pdf

- Van Bruggen, M. (2015). *Communication in the earthquake region in the province of Groningen: An evaluation by the residents* (Ongepubliceerde Masterscriptie). Universiteit Twente.
- Van Iersel, J., Smeulders, E., & Bakker, W. (2014, 12 maart). *Het woon- en leefklimaat in ontwikkeling: Woon- en leefbaarheidsmonitor Eemsdelta*. RIGO Research en Advies BV.
- Vereniging Eigen Huis (2016, november). *Peilstok Groningen*.

## Bijlage 6: Reactie onderzoekers

We hebben in het kader van hoor-wederhoor de onderzoekers van de vijf gereviewde onderzoeken de mogelijkheid geboden een schriftelijke inhoudelijke reactie te geven op ons tweede tussenrapport. Ze konden hierin bijvoorbeeld extra informatie of uitleg geven over de eigen onderzoeks aanpak, reageren op de bevindingen en conclusies van ons review, of anderszins een reactie geven. De afspraak hierbij was dat we de reactie integraal zouden opnemen in het eindrapport. Eén van de vijf betrokken onderzoeksteams heeft voor de mogelijkheid gekozen. We geven deze reactie hieronder integraal weer.

### Reactie

Reactie Gronings Perspectief, Rijksuniversiteit Groningen

Katherine Stroebe & Tom Postmes

Juni 2017

We zijn onder de indruk van de manier waarop onze collega's dit onderzoek hebben uitgevoerd. Uiteraard zijn we tevreden over de manier waarop ons werk wordt beoordeeld: het doet ons deugd dat gerenommeerde collega-onderzoekers van mening zijn dat ons onderzoek een blauwdruk kan zijn voor toekomstig onderzoek en dat het Lifelines onderdeel daarvan als "gouden standaard" gezien kan worden.

In het rapport worden aanbevelingen gedaan die we onderschrijven en waarvan we hopen dat ze de leidraad zijn voor toekomstig onderzoek. We onderschrijven van ganser harte de oproep tot integratie van onderzoek. Wij zijn voorstander van een gecoördineerde aanpak waarin belangrijke thema's (gezondheid, veiligheid, woontevredenheid, leefbaarheid in brede zin) longitudinaal onderzocht worden in relatie tot elkaar (en dus niet gefragmenteerd).

De auteurs van het rapport wijzen er terecht op dat het doen van gedegen onderzoek op de manier zoals zij dat voor zich zien een aanzienlijke investering zal vergen. Daaraan willen wij geen afbreuk doen, maar wij hebben wel een kleine kanttekening, die ons relevant lijkt voor de vormgeving van toekomstig onderzoek. Het belang van een hoge respons wordt wat ons betreft iets te sterk benadrukt.

De onderzoekers wekken de indruk dat een lage respons leidt tot vertekening van resultaten. Dat is discutabel: een lage respons betekent dat je minder zeker bent over de validiteit en generaliseerbaarheid van resultaten. Onzekerheid is niet hetzelfde als vertekening. Wij wijzen erop dat er in Gronings Perspectief geen enkele aanwijzing is dat non-respons systematisch was - met andere woorden, wij vermoeden dat de vertekening gering is. Bovendien kan de onzekerheid die een lage respons met zich meebrengt op meerdere manieren ondervangen worden dan enkel door het verhogen van de respons. In ons onderzoek hebben wij de keuze gemaakt om via Lifelines data te verzamelen om zo de resultaten uit ons panel te "repliceren". Dat leek ons zowel om methodische als budgettaire redenen een goede oplossing. Ons advies is dan ook: zoek ook in toekomstig onderzoek naar mogelijkheden om de resultaten uit een longitudinaal onderzoek te repliceren of valideren. Uiteindelijk is dat het meest probate middel om zekerheid te verschaffen.

Desalniettemin zijn we het volkomen met de onderzoekers eens dat het goed is als de respons in vervolgonderzoek hoger is. Ook zijn we van mening dat de uitval in ons panel het nodig maakt om de steekproef aan te vullen. Tenslotte waarderen we het

bijzonder dat zij erop wijzen dat het vervolgonderzoek de leefbaarheid in brede zin moet monitoren, waardoor verschillende aspecten daarvan in relatie kunnen worden gebracht.

Kortom: wij danken de onderzoekers voor hun waardevolle werk en voor de validatie. We hopen dat het Ministerie van Economische Zaken hun aanbevelingen ter harte neemt. Er liggen prangende inhoudelijke vragen over de maatschappelijke gevolgen van gaswinning. De huidige validatiestudie bevestigt in welke vorm die beantwoord moeten worden.

Katherine Stroebe & Tom Postmes



Universiteit Leiden, afdeling Sociale en Organisatie Psychologie  
Wassenaarseweg 52 T 071 527 4075  
Postbus 9555 E-mail: [emors@fsw.leidenuniv.nl](mailto:emors@fsw.leidenuniv.nl)  
2300 RB Leiden

Universiteit Utrecht, afdeling Methodologie en Statistiek  
Padualaan 14 T 030 253 4438  
Postbus 80.140 E-mail: [p.lugtig@uu.nl](mailto:p.lugtig@uu.nl)  
3508 TC Utrecht

**Eindrapport**