

**Discussiepaper 'de maatschappelijke waarde van kortere en
betrouwbaardere reistijden'**

**Auteur: Niek Mouter
Organisatie: Technische Universiteit Delft
Mailadres: n.mouter@gmail.com**

**"Bijdrage aan het Colloquium Vervoersplanologisch Speurwerk 2013
21 en 22 november, Rotterdam"**

Discussiepaper 'de maatschappelijke waarde van kortere en betrouwbaardere reistijden'

Samenvatting

Het rapport 'de maatschappelijke waarde van kortere en betrouwbaardere reistijden' van het Kennisinstituut voor Mobiliteitsbeleid (KiM) levert nieuwe kengetallen op voor de waardering van de verandering in de gemiddelde reistijd en betrouwbaarheid van die reistijd. De studie is gebaseerd op een achtergrondstudie, uitgevoerd door een consortium onder leiding van Significance.

Het hoofddoel van dit paper is het voeden en scherpstellen van de discussie over het rapport van het KiM en de achtergrondstudie. Het doel van dit paper is niet om de suggestie te wekken dat de kwaliteit van de studie onder de maat is, want dit is mijn mening ook niet. Waarom is het dan zo nuttig om de mitsen en maren van deze studie te bespreken? De hoofdreden is dat 'de maatschappelijke waarde van kortere en betrouwbaardere reistijden' een zeer belangrijke studie is voor de beoordeling van infrastructuurprojecten in de Maatschappelijke kosten-batenanalyse (MKBA). Het kennen van de beperkingen en de sterke punten van een studie zorgt ervoor dat deze beter in de praktijk kan worden toegepast. Als men de beperkingen van een studie niet kent, bestaat het gevaar dat de studie klakkeloos wordt toegepast in situaties waarin het niet verstandig is om deze toe te passen. Als men de sterke punten niet kent, bestaat het gevaar dat men zich alleen richt op de beperkingen en vervolgens het kind met het badwater weggooit. Aangezien de verschillende rapporten over de studie 'maatschappelijke waarde van kortere en betrouwbaardere reistijden' meer aandacht besteden aan de sterke punten dan aan de beperkingen, denk ik dat het eerste gevaar groter is dan het tweede gevaar. Om deze reden richt dit paper zich vooral op de beperkingen van de studie, met name door claims in twijfel te trekken. In het discussiepaper verdedig ik onder meer de volgende stellingen:

- De claim van de KiM-studie dat de uitkomsten representatief zijn voor Nederland, is slecht onderbouwd;
- De claim van de KiM-studie dat de methoden die zijn toegepast state-of-the-art zijn, is incorrect;
- De claim van de KiM-studie dat de uitkomsten in lijn zijn met internationale waarderingstudies, is incorrect;
- Doordat de gevonden kengetallen voor reistijdwaardering en betrouwbaarheidswaardering zijn omgeven met onzekerheden, is het suboptimaal dat de KiM-studie louter puntschattingen communiceert;
- Betrouwbaarheidseffecten van wegprojecten kunnen in een MKBA veel makkelijker worden meegenomen dan betrouwbaarheidseffecten van OV-projecten. Er is sprake van oneerlijke concurrentie. Men moet deze onevenwichtigheid zo snel mogelijk herstellen. Om dit evenwicht te herstellen moet men investeren in het meenemen van betrouwbaarheidseffecten van OV-projecten in een MKBA;
- De KiM-studie levert kengetallen op voor reistijdwaardering van pleziervaarders, maar besteedt geen aandacht aan de reistijdwaardering van fietsers. Gezien de bijdrage van de fiets aan de mobiliteit van Nederlanders is dit bevreemdend.

1. Introductie

Dit paper bespreekt de studie 'de maatschappelijke waarde van kortere en betrouwbaardere reistijden'¹ van het Kennisinstituut voor Mobiliteitsbeleid (KiM). Deze studie van het KiM levert nieuwe kengetallen op voor de waardering van de verandering in de gemiddelde reistijd en betrouwbaarheid van die reistijd en is gebaseerd op de studie 'Values of time and reliability in passenger and freight transport in the Netherlands' (vanaf nu: VoRVoR-studie), uitgevoerd door een consortium onder leiding van Significance.²

Het hoofddoel van dit paper is het voeden en scherpstellen van de discussie over het rapport van het KiM en de achtergrondstudie, omdat het een zeer belangrijke studie is voor de beoordeling van infrastructuurprojecten in de Maatschappelijke kosten-batenanalyse (MKBA). Ook bespreekt het paper een aantal richtingen voor verder onderzoek. In het discussiepaper verdedig ik onder meer de volgende stellingen:

- De claim van de KiM-studie dat de uitkomsten representatief zijn voor Nederland, is slecht onderbouwd;
- De claim van de KiM-studie dat de methoden die zijn toegepast state-of-the-art zijn, is incorrect;
- De claim van de KiM-studie dat de uitkomsten in lijn zijn met internationale waarderingstudies, is incorrect;
- Doordat de gevonden kengetallen voor reistijdwaardering en betrouwbaarheidswaardering zijn omgeven met onzekerheden, is het suboptimaal dat de KiM-studie louter puntschattingen communiceert;
- Betrouwbaarheidseffecten van wegprojecten kunnen in een MKBA veel makkelijker worden meegenomen dan betrouwbaarheidseffecten van OV-projecten. Er is sprake van oneerlijke concurrentie. Men moet deze onevenwichtigheid zo snel mogelijk herstellen. Om dit evenwicht te herstellen moet men investeren in het meenemen van betrouwbaarheidseffecten van OV-projecten in een MKBA;
- De KiM-studie levert kengetallen op voor reistijdwaardering van pleziervaarders, maar besteedt geen aandacht aan de reistijdwaardering van fietsers. Gezien de bijdrage van de fiets aan de mobiliteit van Nederlanders is dit bevreemdend.

Het doel van dit paper is niet om de suggestie te wekken dat de kwaliteit van de studie onder de maat is, want dit is mijn mening ook niet. Waarom dan toch de mitsen en maren bespreken? De hoofdreden is dat het kennen van de beperkingen en de sterke punten van een studie ervoor zorgt dat deze beter in de praktijk kan worden toegepast. Als gebruikers (bijvoorbeeld consultants, beleidsambtenaren, burgers, politici en vertegenwoordigers van belangengroepen) de beperkingen van een studie niet kennen, bestaat het gevaar dat de studie klakkeloos wordt toegepast in situaties waarin het niet verstandig is om deze toe te passen. Als gebruikers de sterke punten niet kennen, bestaat het gevaar dat zij zich alleen richten op de beperkingen en vervolgens het kind met het badwater weggooien. Aangezien de verschillende rapporten over de studie 'maatschappelijke waarde van kortere en betrouwbaardere reistijden' meer aandacht besteden aan de sterke punten³ dan aan de beperkingen, denk ik dat het eerste gevaar groter is dan het tweede gevaar. Het eerste gevaar (beperkingen zijn onbekend) wordt vergroot, doordat de toekomstige gebruiker van de kengetallen relatief diep moet graven om de beperkingen van de studie zelf te kunnen beoordelen. De kengetallen voor reistijdwaardering en betrouwbaarheidswaardering staan straks op de website van het Steunpunt Economische

¹ <http://www.kimnet.nl/sites/kimnet.nl/files/de-maatschappelijke-waarde-van-kortere-en-betrouwbaardere-reistijden.pdf>

² http://www.kimnet.nl/sites/kimnet.nl/files/filemanager/Values_of_time_and_reliability_in_passenger_and_freight_transport_in_The_Netherlands.pdf

³ Voordelen zijn bijvoorbeeld dat een duidelijke stap vooruit is gemaakt ten opzichte van de vorige VoT en VoR studie (Hague Consulting Group, 1998), het rapport is gereviewed door binnenlandse- en buitenlandse experts en voor het eerst zijn nu op praktijkonderzoek gebaseerde kengetallen beschikbaar voor de economische en maatschappelijke waardering van een grotere betrouwbaarheid van reistijden. Voor een geheel overzicht van de voordelen zie KiM (2013) en Significance et al., (2012).

Evaluatie,⁴ voor meer informatie wordt men doorverwezen naar het rapport van het Kennisinstituut voor Mobiliteitsbeleid (KiM, 2013), waarin men verder wordt doorverwezen naar de achtergrondstudie (Significance et al., 2012, een Engelstalig-rapport met relatief veel jargon) voor meer informatie. Ook deze achtergrondstudie bespreekt niet alle beperkingen van de studie. Om deze reden richt ik mij in dit paper vooral op de beperkingen van het onderzoek.

Ter onderbouwing van de kritische beschouwing van de studie 'maatschappelijke waarde van kortere en betrouwbaardere reistijden' zijn de brondocumenten bestudeerd. Alle vragen die bij het lezen van de studies naar boven kwamen werden genoteerd. Deze vragen werden vervolgens met drie onderzoekers die de VoTVoR-studie hebben uitgevoerd of begeleid (Pim Warffemius, Gerard de Jong en Erik Verhoef) besproken. De interviews zorgden ervoor dat een aantal vragen bij de VoTVoR-studie werden weggenomen.⁵ De vragen die niet (volledig) werden weggenomen zullen worden besproken in dit paper. Voor aanvang aan het discussiepaper nog vier opmerkingen ter positionering van het paper:

- 1 Ik stel het erg op prijs dat de drie onderzoekers tijd hebben vrijgemaakt voor een interview, waarin zij louter kritische vragen over hun onderzoek kregen voorgeschoteld;
- 2 De drie onderzoekers waren het met een aantal stellingen die ik poneerde eens, maar met een aantal juist zeer oneens. De standpunten die dit paper inneemt zijn dus geheel voor rekening van de auteur;
- 3 Het doel van het paper is het voeden en scherpstellen van de discussie over tijdswaardering. Om deze reden zijn op enkele plekken met opzet nuances weggelaten;
- 4 De stellingen die ik in dit paper verdedig, zijn vaak net iets scherper dan mijn eigen mening. Waarom? Ik denk dat je als wetenschapper – naast het publiceren in topjournals en het dissemineren van onderzoeksresultaten – soms de rol van de Advocaat van de Duivel moet aannemen om discussies over beleid of beleidsstudies aan te wakkeren (Huys noemt dit de rol de Middeleeuwse nar, 2011). Zoals een advocaat soms een cliënt verdedigt die hij als rechter zou veroordelen, zo reflecteer ik in dit paper kritisch op een studie, die ik wellicht als beleidsambtenaar zonder vragen zou toepassen. Het CVS is denk ik de ideale plaats voor wetenschappers om de rol van Advocaat van de Duivel ten opzichte van het beleid aan te nemen (zie ook, te Brömmelstroet, at al., 2013). Maar dit geldt ook andersom. Wat heeft het beleid gehad aan alle wetenschappelijke studies die de laatste 40 jaar op het snijvlak van verkeer, vervoer en ruimtelijke ordening zijn uitgevoerd?

2. De VoTVoR-studie in een notendop

Het verbeteren van de gemiddelde reistijd en de betrouwbaarheid zijn belangrijke baten van investeringen in transportinfrastructuur (zoals spoorverbreding, nieuwe snelwegen, uitbreiding van een luchthaven of vergroting van een sluis). Deze baten zijn steeds samengesteld uit een 'financiële waardering per eenheid' (P) vermenigvuldigd met een 'hoeveelheid' (Q). Als bijvoorbeeld door de aanleg van een nieuwe weg een fileknelpunt wordt opgelost waardoor reizigers 'Q' uren aan reistijd per jaar besparen, dan zijn de reistijdbaten op jaarbasis 'P' X 'Q'. De VoTVoR-studie gaat over de 'financiële waardering per eenheid (de P)'.

De kengetallen voor de waardering van de verandering in de gemiddelde reistijd (Value of Time, vanaf nu: VoT) en betrouwbaarheid (Value of Reliability, vanaf nu: VoR) kunnen worden toegepast in Maatschappelijke kosten-batenanalyses (MKBA's) voor de waardering van verandering in de gemiddelde reistijd en betrouwbaarheid als gevolg van nieuwe infrastructuurprojecten.⁶

⁴ http://www.rijkswaterstaat.nl/zakelijk/economische_evaluatie/steunpunt_economische_evaluatie/

⁵ Voorbeelden van vragen die tijdens de interviews werden weggenomen zijn: In hoeverre is het verantwoord om mensen met een OV-jaarkaart, mensen met een jaarabonnement en mensen zonder abonnement hetzelfde experiment te laten doen? In hoeverre is er gecorrigeerd voor strategisch antwoordgedrag? In hoeverre kan het verschil in waardering tussen lange en korte reisafstanden worden verklaard door het 'reference point effect' Hjorth en Fosgerau (2012)?

⁶ Merk op dat er kritiek is op het inschatten van baten en kosten van ruimtelijk-infrastructurele projecten aan de hand van reistijdveranderingen en betrouwbaarheidsverandering. Zie bijvoorbeeld, Accessibility is gold, mobility is not: a proposal for the improvement of Dutch transport-related cost-benefits analysis (Ferreira et al., 2012).

Vanaf 2005 neemt men betrouwbaarheidswinsten al mee in MKBA's voor investeringen in wegprojecten. Een internationale expert meeting stelde toen een vuistregel vast voor het meenemen van betrouwbaarheidswinsten. Bij wegprojecten waar sprake is van congestie mocht voor het berekenen van de betrouwbaarheidswinsten een opslag op de reistijdwinsten worden gehanteerd. Er geldt: betrouwbaarheidsbaten = reistijdbaten $\times 0,25$. Dit opslagpercentage van 25% is ontleend aan CPB (2004). Deze opslag van 25% wordt door hoofdrolspelers in de Nederlandse beoordelingspraktijk voor ruimtelijk-infrastructurele projecten als problematisch ervaren (Mouter et al., 2013). Omdat de gebruikte opslag van 25% een grof gemiddelde is, kunnen betrouwbaarheidsbaten voor specifieke betrouwbaarheidsprojecten worden over- of onderschat. Daarnaast laat de opslagmethode alleen betrouwbaarheidsbaten zien als er ook reistijdwinsten zijn. Al in 2005 werd de behoefte geformuleerd om waarderingsskengetallen voor betrouwbaarheidswinsten te formuleren op basis van empirisch onderzoek. Deze behoefte is nu middels de VoTVoR-studie vervuld.

Om de maatschappelijke waardering van reistijd en betrouwbaarheid van reistijd te bepalen, is gebruik gemaakt van zogenoemd stated-preference-onderzoek, waarbij de respondenten situaties krijgen voorgelegd waarin de kosten van de reis, de reistijd en de betrouwbaarheid van de reistijd variëren. Uit de keuzes die respondenten maken, is afgeleid hoe zij reistijd, betrouwbaarheid van de reistijd en geld tegen elkaar afwegen. Het voordeel van stated-preference-onderzoek is dat de onderzoeker controle heeft over de variaties in kosten, reistijd en spreiding van reistijd die aan de respondenten worden voorgelegd.

De dataverzameling voor het personenverkeer en -vervoer is in twee stappen verlopen. Bij de eerste steekproef (2009) is gewerkt met een internetenquête (aantal respondenten: 5.760). Initiële modellen die op deze data werden geschat lieten veel kleinere reistijdwaarderingen zien dan eerdere waarden. Om deze reden is in 2011 nieuwe data verzameld. Reizigers op benzinstations, parkeerplaatsen langs de snelweg en in de stad, treinstations en bushaltes werd gevraagd om deel te nemen aan het onderzoek. Respondenten die mee wilden doen, kregen een weblink naar de vragenlijst mee naar huis (aantal respondenten: 1.430).

Inmiddels zijn er verschillende rapporten en papers over de resultaten van de VoTVoR-studie verschenen. Het paper dat op het CVS gepresenteerd wordt, trekt onder meer de volgende conclusies:

(Warffemius et al., november 2013)

- **Blz. 7: Representatief:** *"In dit hoofdstuk staan de nieuwe waarderingsskengetallen voor veranderingen in de gemiddelde reistijd en de betrouwbaarheid van de reistijd. Deze getallen zijn representatief voor Nederland";*
- **Blz. 7: State-of-the-art:** *"in dit onderzoek zijn de laatste relevante nationale en internationale wetenschappelijke ontwikkelingen verwerkt";*
- **Blz. 8: In lijn met internationale waarderingssstudies:** *"Vergelijking met recente internationale waarderingssstudies toont aan dat de nieuwe Nederlandse VoTs en VoRs uit deze studie goed in lijn zijn met de reistijd- en betrouwbaarheidswaarderingen die in vergelijkbare landen worden gebruikt."*

Op elk van de statements reflecteer ik kritisch verderop in het paper.

3. Kritische beschouwing op de statements die worden gemaakt in de VoTVoR-studie

3.1. De nieuwe VoTs en VoRs zijn representatief voor Nederland.

Warffemius et al. (2013) en KiM (2013) stellen dat de gevonden kengetallen in de VoTVoR-studie representatief zijn voor Nederland. In andere woorden, de mensen die hebben deelgenomen aan het onderzoek zijn een goede afspiegeling van de populatie Nederlandse reizigers (zie CBS: Onderzoek Verplaatsingen in Nederland

2010, afgekort OViN, voor informatie over deze populatie). En, je kunt op basis van deze steekproef uitspraken doen over de VoT en de VoR van de gemiddelde Nederlandse reiziger gerelateerd aan het kenmerk van de reis. Ik beargumenteer dat deze claim slecht onderbouwd (en waarschijnlijk incorrect) is om een aantal redenen. Voordat ik deze redenen bespreek, eerst wat uitleg over het begrip 'representativiteit' in relatie tot deze studie.

3.1.1. Wanneer is er sprake van representativiteit?

Er is pas sprake van een representatieve steekproef als de specifieke persoonskenmerken die bepalend zijn voor waardering van reistijd en betrouwbaarheid in dezelfde verhouding voorkomen in de steekproef en de populatie. Meer specifiek, als 'geslacht' invloed heeft op de waardering van reistijd dan wil je idealiter dat de percentages mannen en vrouwen in de steekproef overeenkomen met de percentages mannen en vrouwen in de populatie. De steekproef moet zoveel mogelijk een miniatuur zijn van de populatie.

De VoTVoR-studie onderscheidt vijf specifieke persoonskenmerken die – volgens de literatuur – bepalend kunnen zijn voor de waardering van reistijd en betrouwbaarheid. Naast 'geslacht' zijn dit 'leeftijd', 'educatie', 'inkomen' en 'type huishouden'. Als de steekproef representatief is voor deze kenmerken, dan is de steekproef ook representatief voor de reistijdwaardering van de Nederlandse reiziger. Vreemd genoeg beschrijft de studie van Significance et al., (2012) niet in welke verhouding de verschillende kenmerken in de steekproef en in de populatie vertegenwoordigd zijn. Een tabel die de deelnemers aan het onderzoek uitsplitst naar de vijf relevante persoonskenmerken en een tabel met een uitsnede van Nederlandse reizigers naar deze vijf kenmerken (OViN) zou van nut zijn geweest. Hierdoor zou het mogelijk zijn voor de lezer om de representativiteit van de studie zelf te beoordelen.

Hoewel men in de achtergrondstudie (Significance et al., 2012) – waarop de conclusies van Warffemius et al. (2013) en KiM (2013) zijn gebaseerd – niet expliciet de conclusie trekt dat de steekproef representatief is voor Nederland,⁷ stelt men wel dat de studie representatief is gemaakt door weging toe te passen (Significance et al., 2012, blz. 112):

“In the second step, we make our survey representative for the mobility of the Dutch population. For this, we divided all trips in the OViN (Onderzoek Verplaatsingen in Nederland 2010 by Statistics Netherlands) based on five population variables (gender, age, income, household composition and education) and two trip variables (period of the day combined with travel mode and travel time category combined with travel mode)....The distribution of the trips in our survey over the seven variables is different from the whole Dutch population (e.g. many more commuters are present in the VOTVOR survey than one would expect in a typical sample). An Iterative Proportional Fitting method was used to calculate new weights for our VOTVOR survey such that the weighted distributions for the seven variables match the weighted distributions of the OViN survey. In calculating the final values of time, the value of time for each respondent is weighted with this weight factor as determined by this expansion procedure, and with the travel time.”

Ik stel dat het onzeker is of de stelling 'we make our survey representative for the mobility of the Dutch population' klopt. Indien een relevante subgroep⁸ in de steekproef ondervertegenwoordigd is, kan men via weging ervoor zorgen dat gecorrigeerd wordt voor niet-representativiteit bij het schatten van de VoT en de VoR. Door weging toe te passen, kun je op basis van een niet-representatieve steekproef toch representatieve uitspraken doen. Een noodzakelijke voorwaarde is wel dat alle relevante subgroepen binnen de populatie vertegenwoordigd zijn in de steekproef. Wegen zorgt er niet ineens voor dat je observaties hebt van mensen waarvan je geen observaties hebt. Meer specifiek, als je een populatie hebt van 500 mannen en 500 vrouwen

⁷ Het doel van dit project was om de officiële waarden voor tijd in zowel het personenverkeer en- vervoer als het goederenvervoer te actualiseren en niet zozeer om representatieve waarden te genereren.

⁸ Voorbeeld van een subgroep: hoogopgeleide alleenstaande vrouwen onder de 35 jaar, met een hoog inkomen.

en je trekt een steekproef van 100 mannen en 0 vrouwen, dan helpt weging niet om met deze niet-representatieve steekproef een representatieve uitspraak te doen over de populatie, omdat je geen observaties hebt van vrouwen. Een tabel die uitwijst dat alle subgroepen vertegenwoordigd zijn in de steekproef, zou helpen om te kunnen bepalen of aan de voorwaarde ‘van alle relevante subgroepen zijn er observaties’ voldaan is. Ben je er dan als je van elke relevante subgroep ten minste één observatie hebt? Dit is een vraag die niet ondubbelzinnig kan worden beantwoord. Indien in het bovenstaande voorbeeld de steekproef bestaat uit 2 vrouwen en 98 mannen, bestaat er een kans dat de reistijdwaardering van deze 2 vrouwen sterk afwijkt van de andere 498 vrouwen in de populatie. De reistijdwaardering van deze 2 vrouwen geeft dan een sterk vertekend beeld. Een weging van de steekproef van 2 vrouwen 98 mannen zorgt er niet voor dat representatieve uitspraken kunnen worden gedaan voor de populatie. Het risico op een vertekend beeld binnen een subgroep neemt af als het aantal respondenten per subgroep toeneemt. Bij een steekproef van 10 vrouwen en 90 mannen wordt het risico bijvoorbeeld kleiner. Kortom, de uitspraak dat men met wegingen de uitkomsten van de steekproef representatief maakt voor de gemiddelde Nederlandse reiziger kan alleen worden bevestigd als duidelijk is dat men van elke subgroep voldoende observaties heeft. Nu het rapport geen informatie geeft om te kunnen controleren of aan deze voorwaarde is voldaan, is het onzeker of de uitspraak klopt dat de nieuwe waarderingen voor reistijd en reistijdbetrouwbaarheid representatief is gemaakt voor Nederland door wegingen toe te passen. Nogmaals, het is raadzaam om informatie over de steekproef en de populatie in het KiM-rapport of de achtergrondstudie te presenteren, zodat gebruikers van de studie zelf een oordeel kunnen vellen over de representativiteit.

3.1.2. Representativiteit bus/tram/metro, luchtvaart en pleziervaart moeilijk te verdedigen

Ik stel dat het voor drie groepen modaliteiten moeilijk te verdedigen is dat er – ondanks het toepassen van wegingen – representatieve uitspraken kunnen worden gedaan. Ten eerste de groep zakelijk verkeer bus/tram/metro. De VoT en de VoR zijn vastgesteld op basis van observaties van slechts 11 respondenten. Indien je een VoT en een VoR schat op basis van keuzes van 11 respondenten is het risico groot dat een relevante subgroep niet in de steekproef zit en/of dat personen met een afwijkende reistijdwaardering de resultaten sterk beïnvloeden (zie 3.1.4). Het gevaar bestaat dat de VoT en de VoR geen goede weergave zijn van de werkelijke VoT en VoR. Dit geldt ook voor de gecombineerde VoT en VoR voor ‘luchtvaart woon-werk’ + ‘luchtvaart zakelijk’. Deze is gebaseerd op slechts 30 respondenten (Significance, 2012; tabel 38, blz. 83). Stel dat er een MKBA komt van een nieuwe Kaagbaan, dan worden betrouwbaarheidswinsten en reistijdwinsten voor reizigers met het motief ‘zakelijk’ en ‘woon-werk’ vastgesteld op basis van observaties van slechts 30 mensen.

De derde groep VoTs waarbij representativiteit lastig te verdedigen lijkt te zijn is de VoT voor pleziervaarders. De respondenten zijn gerekruteerd in het Gooi (Naarden en Muiden) en in Bruinisse (Zeeland). Dit is mogelijk één type pleziervaarders met waarschijnlijk een hoger gemiddeld inkomen (en waardering voor reistijdwinst) dan de gemiddelde pleziervaarder. Je weet absoluut niet zeker of de gevonden VoT representatief is voor Nederlandse pleziervaarders, aangezien andere groepen pleziervaarders, zoals pleziervaarders op de Friese meren niet zijn opgenomen in de steekproef. Mogelijk verschilt het gedrag van deze groep pleziervaarders van de pleziervaarders die in de steekproef zijn opgenomen en heeft deze groep een afwijkende waardering voor verandering in reistijd. Dat men pleziervaarders op de Friese meren niet heeft opgenomen in de steekproef is extra opmerkelijk nu de VoTs voor de pleziervaart juist zullen worden gebruikt voor MKBAs van sluisprojecten en brugprojecten in Friesland. Kortom, hoogstwaarschijnlijk kun je op basis van de VoTVoR-studie een representatieve uitspraak doen over de VoT van pleziervaarders uit het Gooi (of voor pleziervaarders met soortgelijke kenmerken, denk aan Wassenaar), maar niet voor pleziervaarders op de Friese meren. Dit, terwijl de VoT voor de pleziervaart juist voor Friese projecten is bedoeld.

3.1.3. Pace of life

Uit de voorgaande paragrafen volgt dat representativiteit gewaarborgd is als elke relevante subgroep in de populatie voldoende vertegenwoordigd is in de steekproef. Relevante subgroepen bepaalt men door vast te

stellen welke persoonskenmerken de waardering van reistijd en betrouwbaarheid beïnvloeden. De VoTVoR-studie onderscheidt vijf persoonskenmerken te weten: ‘geslacht’, ‘leeftijd’, ‘educatie’, ‘inkomen’ en ‘type huishouden.’ Mijns inziens, zijn dit niet de enige vijf persoonskenmerken die de waardering van reistijd en betrouwbaarheid beïnvloeden. ‘Levensstijl’ is bijvoorbeeld een variabele die invloed heeft op reistijdwaardering. Dit punt illustreer ik met een uitspraak van een Noorse respondent die ik interviewde voor mijn studie ‘MKBA Internationaal’ (Mouter, forthcoming): *“dat mensen in Oslo tijd anders waarderen dan in Noord-Noorwegen kun je met het blote oog zien. Mensen in Oslo hebben altijd haast, mensen op het platteland doen het rustig aan. Tuurlijk heeft dit ook te maken met inkomen en educatie, maar het zit ook in de mentaliteit.”* In ‘The geography of time’ werkt psycholoog Robert Levine (2006) uit dat cultuur een grote invloed heeft op het perspectief van tijd van mensen, doordat zij een verschillende ‘Pace of Life’ hebben. Levine vergeleek onder meer het verschil in ‘Pace of Life’ tussen 60 Amerikaanse steden. Philip Zimbardo (bekend van het Stanford prison experiment) werkt in een korte presentatie (<http://www.youtube.com/watch?v=A3oliH7BLmg>) een aantal resultaten van Levine uit en stelt in een eigen onderzoek vast dat Noord-Italianen een andere ‘Pace of Life’ hebben dan Sicilianen.

In analogie met de resultaten van Levine en Zimbardo kun je in Nederland verwachten dat mensen in de Randstad (en misschien in de regio Arnhem-Nijmegen en Brabantstad) een meer gehaaste ‘Pace of Life’ hebben dan mensen in de rest van Nederland. Naar mijn idee is het niet onaannemelijk dat ‘Pace of Life’ (gecorrigeerd voor de vijf andere persoonskenmerken) de waardering voor reistijdwinst beïnvloedt (High Pace of Life, hoge reistijdwaardering). Het is onduidelijk of de VoTVoR-studie 2011 hier expliciet rekening mee houdt. In ieder geval is het duidelijk dat respondenten vooral zijn gerekruteerd op plaatsen waar je veel mensen met een ‘High Pace of Life’ verwacht (Rotterdam Groothandelsgebouw, Amsterdam Muntplein, Den Haag CS, Eindhoven CS, Leiden CS, Den Bosch CS, Utrecht CS, Rotterdam CS, Arnhem CS etc). Ook is het zeker dat er niet is gerekruteerd in provincies als Overijssel, Limburg, Groningen, Friesland en Flevoland waar je mensen met een ‘Low Pace of Life’ verwacht. Indien mensen met een ‘Low Pace of Life’ ondervertegenwoordigd zijn in de steekproef, heeft dit tot gevolg dat de VoTs en VoRs die voortvloeien uit de VoTVoR-studie overschattingen zijn van de daadwerkelijke VoTs en VoRs van de gemiddelde Nederlandse reiziger. Naar mijn idee kan dit probleem op twee manieren worden opgelost:

- Je voegt ‘Pace of Life’ toe als persoonskenmerk. Je probeert ‘Pace of Life’ zo goed mogelijk te operationaliseren. Vervolgens onderzoek je hoe de populatie kan worden onderverdeeld naar ‘Pace of Life’ (hoeveel Nederlandse reizigers hebben High, Medium en Low ‘Pace of Life’?). Je neemt vragen op in de enquête om ‘Pace of Life’ van de respondenten vast te stellen. Ten slotte controleer je of de verschillende ‘Paces of Life’ voldoende vertegenwoordigd zijn in de steekproef in relatie tot de populatie;
- Je rekruteert ook in gebieden waar de verwachting is dat mensen een ‘Low Pace of Life’ hebben. Bijvoorbeeld Overijssel, Limburg, Groningen, Friesland en Flevoland.

De eerste optie is wetenschappelijk gezien eleganter. Voor het herstellen van problemen met de huidige VoTVoR-studie lijkt echter alleen de tweede optie een realistisch alternatief. Je probeert ervoor te zorgen dat je voldoende mensen met een High ‘Pace of Life’ en een Low ‘Pace of Life’ in je steekproef hebt, door zowel te rekruteren in gebieden waar je veel mensen verwacht met een High ‘Pace of Life’ als in gebieden waar je veel mensen verwacht met een Low ‘Pace of Life’.

Aanbevelingen: ik stelt dat het verstandig is om aan de volgende punten aandacht te besteden, indien de studie wordt herhaald of alsnog wordt uitgebreid:

- Meer rekruteren in niet-stedelijke gebieden (Overijssel, Limburg, Groningen, Friesland en Flevoland);
- Stel eerst vast wat voor soorten pleziervaarders er allemaal zouden kunnen zijn; bijvoorbeeld: zeilers op Oosterschelde / Westerschelde in Zuid-Holland en Zeeland, mensen die in een sloep varen op de

Friese meren etc. Vervolgens is het verstandig om voor elke type pleziervaarders voldoende respondenten te rekruteren;

- Meer respondenten in het segment 'luchtverkeer zakelijk' + 'luchtverkeer woon-werk';
- Meer respondenten in het segment 'bus/tram/metro zakelijk';
- Neem een tabel op in het rapport die weergeeft hoeveel respondenten er per rekruteringslocatie hebben meegedaan;
- Neem een tabel op in het rapport waarin zowel de steekproef als de populatie wordt uitgesplitst naar de persoonskenmerken die samenhangen met VoT en VoR;
- Voer een non-response analyse uit.

De laatste aanbeveling 'voer een non-response analyse uit' verdient nadere uitleg. Wat is een non-response analyse en wat is het nut ervan?

3.1.4. Non-response analyse

In een non-response analyse onderzoek je of er verschillen bestaan tussen mensen die wel deelnemen aan het onderzoek en mensen die niet deelnemen aan het onderzoek, als het gaat om persoonskenmerken die relevant zijn voor de variabelen die onderwerp zijn van onderzoek (in dit geval de VoT en de VoR). Indien een groep mensen met een hoge reistijdwaardering om wat voor reden dan ook ondervertegenwoordigd is in de steekproef dan bestaat de kans dat de VoTs die uit het onderzoek komen, lager zijn dan de VoTs van de gemiddelde Nederlandse reiziger. Uit de achtergrondstudie volgt dat men de onderzoeksresultaten van de PanelClix-studie in 2009 om deze reden niet vertrouwde (Significance et al., 2012. Blz. 100):

"This means that in the 2009 survey there was a bias towards low-VOT people that have the time to be a member of an internet panel, and an additional bias towards those internet panel members that were willing to give up the time required to actually participate in this survey. Probably, people who cannot afford much (after correction for income) are more likely to participate in an internet panel and to fill in the web questionnaire for a reward. These people have a relatively low VOT."

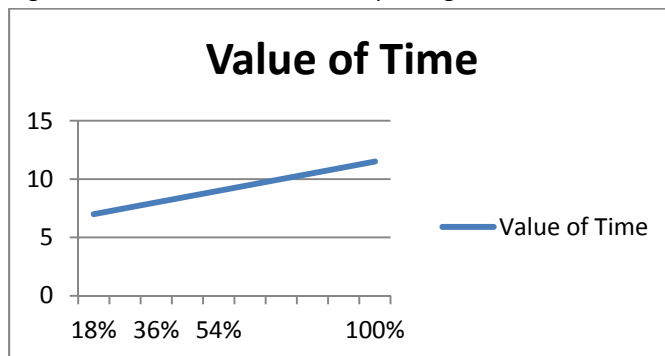
Om deze reden startte men in 2011 een nieuw onderzoek. De vraag is echter hoe men heeft vastgesteld dat de resultaten van het onderzoek in 2011 wel goed zijn? Was hier geen sprake van 'bias towards low-VOT people', of juist een 'bias towards high-VOT people'?⁹ Dit is onbekend. Je zou dit alleen kunnen controleren door gegevens in te winnen van mensen die niet hebben deelgenomen aan het onderzoek en deze te vergelijken met de mensen die wel hebben deelgenomen aan het onderzoek.

Hoe kun je dit organiseren? Potentiele deelnemers zijn voor de VoTVoR 2011-studie gerekruteerd bij een benzinstation of een parkeergarage. Als de deelnemers toestemden mee te doen, schreef de persoon die hen rekruteerde hun mailadres op. Dit was een goed moment om naast het mailadres ook andere gegevens te noteren (bijvoorbeeld inkomen, maar het is waarschijnlijk verstandiger om één vraag uit het keuze-experiment voor te leggen, waarbij de kans groot is dat iemand met een hoge reistijdwaardering voor optie A kiest en iemand met een lage reistijdwaardering voor optie B). Uiteindelijk zijn 3650 personen gerekruteerd, waarvan 1671 personen de enquête uiteindelijk afmaakten. Men zou een non-response analyse kunnen uitvoeren door de gegevens die men noteerde van de 1979 mensen die niet meededen te vergelijken met de gegevens van de 1671 personen die wel meededen. Zeker bij pleziervaart en luchtvaart is het interessant om te analyseren of de non-respons niet alleen bestond uit mensen met een veel hogere VoT dan de mensen die wel hebben deelgenomen aan het onderzoek.

⁹ Een argument voor een 'bias towards high-VOT people', zou kunnen zijn dat mensen die geen toegang hebben tot internet en mensen die weinig gebruik maken van internet, minder kans hebben om deel te nemen aan het onderzoek, omdat gebruik is gemaakt van een webenquête. Als we aannemen dat dit mensen zijn met een lage Value of Time (na correctie voor inkomen), zijn de VoTs en VoRs die volgen uit de VoTVoR-studie overschattingen.

Armstrong en Overton (1977) bespreken een methode waarop je een non-response analyse op een ‘Quick and Dirty’ manier kunt uitvoeren. Armstrong en Overton nemen aan dat mensen die pas na reminders reageren op een enquête meer lijken op de non-response dan de mensen die direct de enquête hebben ingevuld. Zij stellen voor om apart de gevonden uitkomsten van mensen die direct reageren, mensen die reageren na één reminder en mensen die reageren na twee reminders in te schatten, om vervolgens te extrapoleren naar de mensen die – ook na reminders – niet hebben deelgenomen aan het onderzoek (bij de VoTVoR-studie is de non-response 45%). Het idee van Armstrong en Overton illustreer ik hieronder met een fictief voorbeeld. Stel dat 54% van de mensen die zijn gerekruteerd, uiteindelijk hebben meegedaan aan het onderzoek. 18% reageerde direct en heeft een VoT van 7 euro; 18% reageerde na de eerste reminder en heeft een VoT van 8 euro; 18% reageerde pas na de tweede reminder en heeft een VoT van 9 euro. Als je gaat rekenen met de mensen die hebben meegedaan aan het onderzoek vind je waarschijnlijk een VoT van rond de 8 euro. Armstrong en Overton stellen dat dit onjuist is. De kans is aanzienlijk dat een groot deel van de mensen die niet hebben meegedaan aan het onderzoek een hogere VoT hebben dan 9 euro (zie figuur 1). In dit fictieve voorbeeld is er urgentie om nieuwe data te verzamelen. De VoT van de respons, verschilt sterk van de VoT van de non-response, waardoor de VoT wordt onderschat als je alleen met de respons rekent.

Figuur 1: fictief voorbeeld van toepassing van methode Armstrong en Overton (1977) op een VoT-studie



Ik beveel aan om in ieder geval een ‘quick and dirty’ non-response analyse te doen voor de VoTVoR-studie door te extrapoleren vanuit de VoTs van mensen die gelijk de enquête invulden en de mensen die pas na een reminder de enquête invulden.

3.2 De uitkomsten zijn in lijn met internationale waarderingsstudies

KIM (2013, blz. 15) concludeert dat vergelijking met recente internationale waarderingsstudies aantoont dat de nieuwe Nederlandse VoTs en VoRs uit deze studie goed in lijn zijn met de reistijd- en betrouwbaarheidswaarderingen die in vergelijkbare landen worden gebruikt’. De Noorse en Zweedse VoT-studies worden aangehaald als meeste moderne studies in vergelijkbare landen. Deze conclusie gebruikt men ook als argument dat de gevonden VoTs en VoRs uit de VoTVoR-studie 2011 correct zijn (in tegenstelling tot de de VoTs en VoRs die voortvloeiden uit de VoTVoR-studie 2009).

Even een gedachtenexperiment: in zekere zin is hier sprake van confirmation bias. Als de resultaten overeenkomen met resultaten van vergelijkbare landen (VoTVoR-studie 2011) dan kloppen de uitkomsten waarschijnlijk en studeert men niet door. Wijken de resultaten af (VoTVoR-studie 2009), dan is er vast iets mis en studeert men door, totdat men VoTs en VoRs vindt die lijken op de VoTs en VoRs van vergelijkbare landen. Confirmation bias brengt in deze casus twee potentiële problemen met zich mee. Ten eerste, als de Nederlandse VoTs en VoRs in werkelijkheid afwijken van het gemiddelde, dan komt dit nooit als resultaat uit een studie, omdat men doorstudeert totdat men waarden vindt die in de buurt van het gemiddelde zitten. Ten tweede, als je wilt dat de waarden in de buurt zitten van waarden in vergelijkbare landen, zou je kunnen stellen dat de studie zonde van het geld is. Het is goedkoper om gewoon de gemiddelde waarden van vergelijkbare landen als

VoT en VoR te nemen, als op voorhand duidelijk is dat de resultaten van de studie rond het gemiddelde van vergelijkbare landen moet liggen.

Ik stel dat deze conclusie ‘de uitkomsten zijn in lijn met internationale waarderingsstudies’ te stellig is als je de VoT en de VoR als puntschattingen in de MKBA gebruikt. Mijn argument is dat het zeer opmerkelijk is dat de VoT voor treinreizigers in Zweden 7.20 euro is (Significance et al., 2012) en in Zwitserland 20.57 euro (Axhausen et al., 2006, geciteerd in Wardman et al., 2012), tegenover een VoT van 11.50 euro in Nederland (zie Significance et al., blz. 122). Daarbij is de ‘VoT’ voor ‘woon-werk autoverkeer’ in Nederland 9.25 euro. Dit wijkt af van Noorwegen: 12.13 – 26.95 euro (Ramjerdi et al., 2010), Denemarken: 11.87 euro (Fosgerau et al., 2007), Engeland 6.53 euro (Mackie et al., 2003) en Duitsland 5.95 euro (Wardman et al., 2012). De VoT voor ‘zakelijk autoverkeer’ is in Nederland 26.25 euro. Dit wijkt af van Noorwegen 51.20 euro (Ramjerdi et al., 2010). Ik stel dat dit te grote verschillen zijn om te concluderen dat de puntschattingen van de Nederlandse studie in lijn zijn met vergelijkbare landen.

3.2.1 Doe gevoeligheidsanalyses op de VoT en de VoR

Een argument om – ondanks de duidelijke verschillen tussen de puntschattingen van de Nederlandse studie en de buitenlandse studies – te kunnen concluderen dat de nieuwe Nederlandse VoTs en VoRs goed in lijn zijn met de reistijd- en betrouwbaarheidswaarderingen van de meest moderne studies uit vergelijkbare landen, is dat de puntschattingen van de Noorse en de Zweedse studie binnen de bandbreedte liggen van de Nederlandse VoTs. De gevonden VoT voor wegverkeer van 9.25 euro per uur is geen kiezelhard getal. De standaarddeviatie van de gevonden VoT voor wegverkeer (9.25 euro) is ongeveer 2.7. Dit betekent dat bij herhaling van de VoT/VoR-studie de kans 15.9% is dat de VoT voor wegverkeer boven de 11.95 euro ligt en de kans ook 15.9% is dat de gevonden VoT onder de 6.55 euro ligt. In het KiM-rapport staat echter geen zinnig opgenomen als: *“de gevonden VoT voor wegverkeer is 9.25 euro. Indien het onderzoek wordt herhaald, bestaat de kans dat wij een VoT van boven de 12 euro vinden of van onder de 7 euro, daarom bevelen wij aan om in een MKBA altijd een gevoeligheidsanalyse te doen voor een VoT van 12 euro en een VoT van 7 euro.”* Ik stel dat het – gezien de onzekerheden rondom de VoT en de VoR – belangrijk is om gevoeligheidsanalyses op de VoT en de VoR verplicht te stellen in MKBAs

3.3 De studie is state-of-the-art

De uitvoerders van de VoT/VoR-studie zijn zorgvuldig te werk gegaan als het gaat om het inrichten van het onderzoeksproces rondom data-analysemethoden die zijn gebruikt. De wetenschappelijke kwaliteit is steeds geborgd door nauw samen te werken met nationale en internationale experts en wetenschappers. Het is echter de vraag of het ontwerp van het keuze-experiment en de dataverzameling state-of-the-art is geweest. Eerder behandelde ik de vraag of de manier van rekruteren ervoor heeft gezorgd dat representatieve uitspraken over de VoT en de VoR voor Nederlandse reizigers kunnen worden gedaan. Ik stel dat er pas sprake is van een state-of-the-art onderzoek als er meer aandacht zou zijn besteed aan de representativiteit, door te rekruteren onder meer groepen dan nu het geval is en door het uitvoeren van een goede non-response analyse. Ook beargumenteer ik dat de documentatie niet state-of-the-art is. Men had inzicht moeten geven in de onderverdeling van respondenten naar persoonskenmerk (zoals geslacht, leeftijd en inkomen) en naar rekruteringslocatie (hoeveel pleziervaarders in Bruinisse, Naarden en Muiden?). Dit laatste punt is op zich nog gemakkelijk te herstellen. Hieronder behandel ik twee andere voorwaarden waar – zo betoog ik – aan voldaan zou moeten worden om de studie state-of-the-art te noemen. Deze voorwaarden richten zich meer op het ontwerp van het keuze-experiment.

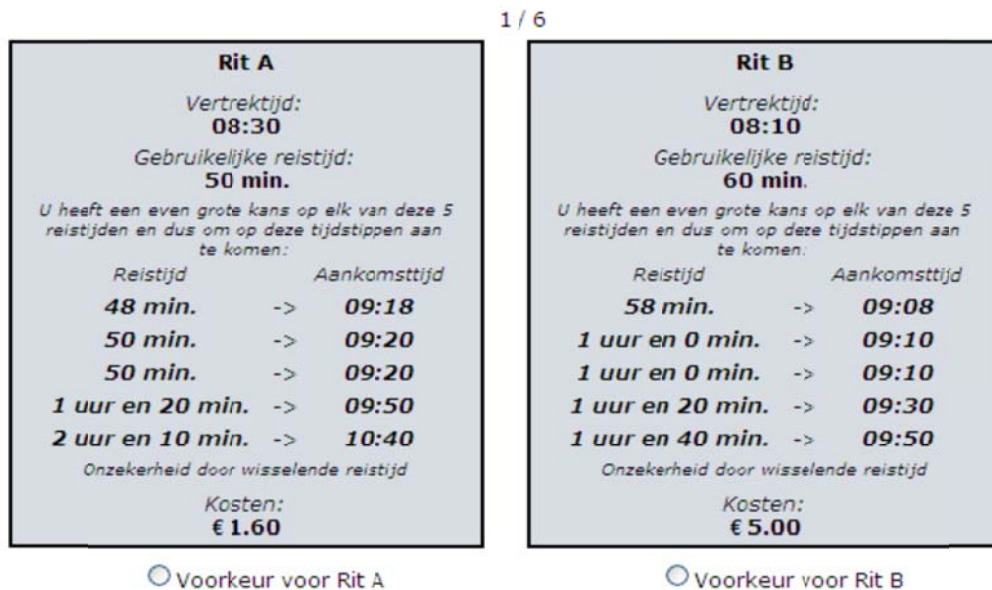
3.3.1 In hoeverre zijn de keuzesituaties realistisch genoeg?

Voor het schatten van de VoT en de VoR in een model moet je de attributen (kosten, betrouwbaarheid en reistijd) in het stated-preference-onderzoek variëren. Door deze variatie komt het voor dat keuzesituaties die men voorgelegd krijgt, enigszins afwijken van de keuzesituaties die men in werkelijkheid meemaakt. Het verschil in opties moet duidelijk voelbaar zijn voor de respondent. Het is echter belangrijk dat de keuzesituaties

niet zover zijn losgezongen van de realiteit, omdat het anders de vraag is in hoeverre respondenten nog een serieus antwoord geven. De motivatie van respondenten die deelnemen aan keuze-experimenten – waarbij de beloning niet afhangt van de kwaliteit van de antwoorden die worden gegeven – mag niet worden overschat. Ik stel dat de keuzesituaties van de VoTVoR-studie veel vraagt van het voorstellingsvermogen van respondenten. Wellicht te veel. De volgende keuzesituatie wordt op blz. 67 van het Significance rapport (2012) gegeven als voorbeeld voor treinreizigers die hebben deelgenomen:

Figuur 2: Voorbeeld van een keuzesituatie voor treinreizigers

Welke situatie heeft uw voorkeur?



Ik zie twee punten van kritiek op het niveau van realisme van deze keuzesituatie.

Opt-out

Ten eerste heb je in de werkelijkheid bij hooguit 5% van de treinreizen die je maakt meer dan een kwartier vertraging. In dit keuze-experiment, heb je in beide situaties 40% kans op meer dan 20 minuten vertraging en bij keuze A heb je zelfs 20% kans op 80 minuten vertraging. Veel mensen kiezen er in werkelijkheid waarschijnlijk voor om deze reis helemaal niet te maken. Zij blijven thuis of kiezen voor een ander vervoersmiddel. Het gevaar bestaat dat respondenten de bovenstaande keuzesituatie niet serieus genoeg vinden, waardoor zij niet serieus nadenken over hun voorkeur bij het invullen van de keuze.

Hoe zou je dit keuze-experiment wel state-of-the-art kunnen maken? Het keuze-experiment zou state-of-the-art zijn, als men (alle of de helft van de) respondenten een opt-out optie zou bieden.¹⁰ Naast alternatief A en B kan de respondent ervoor kiezen om de reis niet te maken. Als de respondent voor de opt-out kiest, moet hij daarna alsnog kiezen voor alternatief A of B. Vervolgens schat je een model inclusief en exclusief de keuzesituaties waarin de respondent eerst voor opt-out heeft gekozen. Het gevolg hiervan is dat de enquête meer tijd kost, waardoor minder respondenten de enquête afmaken. Dit kun je ondervangen door per enquête minder vragen te stellen en tegelijkertijd meer respondenten te rekruteren.

Kansenverhouding

Ten tweede stel ik dat het lastig is voor treinreizigers om je voor te stellen dat er vijf verschillende aankomsttijden zijn met elk een kans van 20% (zie figuur 1), terwijl de betrouwbaarheid van een treinreis er in

¹⁰ 'State-of-the-art' definieer ik als het hoogste niveau van kwaliteit dat met de huidige kennis bereikt kan worden.

werkelijkheid ongeveer als volgt uitziet: 'op tijd: 90% kans; kleine vertraging: 5% kans; grote vertraging: 5% kans'. Het argument om respondenten vijf aankomstmogelijkheden met elk 20% kans voor te schotelen in plaats van de realistische kansenverhouding (90%; 5%; 5%), is dat mensen in keuze-experimenten slecht met kleine kansen en verschillende kansen om kunnen gaan (zie onder meer het werk van Daniel Kahneman). Het is echter de vraag of deze algemene conclusie van Kahneman ook op dit onderzoek van toepassing is. Het is naar mijn idee niet onaannemelijk dat respondenten die dagelijks met de trein reizen, vrij goed uit de voeten kunnen met de kansenverhouding 90%/5%/5%. Zeker als je van tevoren aangeeft dat dit een realistische kansenverhouding is.

Het onderzoeksteam heeft in een pilotstudie gekeken welke vorm van presentatie van 'vijf verschillende aankomsttijden met elk een kans van 20%' het best werd begrepen door respondenten (bijvoorbeeld: histogram, verbaal, taartdiagram, percentages). Bijvoorbeeld door verschillende formats aan de respondenten voor te leggen en hen vervolgens te interviewen (Tseng et al., 2009). Dit was een uitgelezen kans geweest om ook te kijken welke attribootverhouding door de respondenten het best werd begrepen (5x 20%; 3x 33%; 90%, 5%, 5%).

Hoe zou je dit keuze-experiment wel state-of-the-art kunnen maken? Het zou elegant zijn geweest als men het keuze-experiment had uitgevoerd met de realistische kansenverhouding (90%, 5%, 5%) naast de huidige kansenverhouding (5x 20%) en voor beide experimenten een model schatten. Ook dit zal leiden tot extra onderzoekslasten. Hoge onderzoekslasten zijn echter onvermijdelijk als 'state-of-the-art' het doel van het onderzoek is.

Kortom: ik betoog dat het de vraag is of het ontwerp van het keuze-experiment en de dataverzameling state-of-the-art is geweest. Had een realistischer ontwerp van het keuze-experiment en een betere dataverzameling tot andere VoTs en VoRs geleid? Deze vraag is alleen te beantwoorden door deze exercitie daadwerkelijk uit te voeren.

4. Overige aanbevelingen reistijdwaardering en betrouwbaarheidswaardering in de MKBA

De VoTVoR-studie levert een aantal kengetallen op voor de waardering van reistijdverandering en betrouwbaarheidsverandering. Welke stappen zouden nog kunnen worden gemaakt in dit onderzoeksgebied? In deze sectie bespreek ik drie belangrijke stappen.

4.1 Betrouwbaarheidswinsten voor OV-projecten

Uit de VoTVoR-studie volgt dat betrouwbaarheidswinsten in de status quo voor de modaliteiten weg en binnenvaart wel kunnen worden meegenomen middels een opslag op de reistijdwinsten. Voor andere modaliteiten trein, bus/tram/metro en vliegtuig is een dergelijke tijdelijke rekenregel niet beschikbaar. Daardoor kunnen betrouwbaarheidswinsten die ontstaan door investeringen in deze infrastructuur, alleen in MKBAs worden meegenomen door deze kwalitatief te beschrijven. Uit het KiM rapport (2013) volgt dat voor andere modaliteiten dan weg of binnenvaart (trein, bus/tram/metro en vliegtuig) goede informatie ontbreekt over de effecten van beleidsmaatregelen op de betrouwbaarheid van reistijden. Daarnaast is een tijdelijke rekenregel – zoals de 25% opslag die bij het wegverkeer wordt gebruikt en de 15% opslag die bij de binnenvaart wordt gebruikt – niet beschikbaar. Daardoor kunnen betrouwbaarheidswinsten die ontstaan door investeringen in deze infrastructuur, niet in de MKBA worden meegenomen.

Wat is het probleem? Uit verschillende studies volgt dat effecten die niet in geld zijn uitgedrukt een relatief zwakke positie in MKBA-rapporten hebben ten opzichte van effecten die wel in geld zijn uitgedrukt, wat kan leiden tot een zwakkere positie voor deze effecten in het besluitvormingsproces (zie bijvoorbeeld Mackie & Preston, 1998; Mishan, 1988; Mouter, 2012). Indien betrouwbaarheidswinsten bij wegprojecten wel in het MKBA-saldo worden meegenomen en bij OV-projecten niet, dan is er sprake van oneerlijke concurrentie. Zie in figuur 3 een versimpelde MKBA-tabel voor een wegproject en een OV-project met hetzelfde welvaartseffect:

Figuur 3: fictief voorbeeld van een versimpelde MKBA voor een wegproject en een OV-project

Wegproject A		OV-project B	
Reistijdwinst	60.000 euro	Reistijdwinsten	60.000 euro
Betrouwbaarheidswinst	15.000 euro	Betrouwbaarheidswinst	Bij dit project treden betrouwbaarheidswinsten op
Constructiekosten	70.000 euro	Constructiekosten	70.000 euro
Totaal	+5.000 euro	Totaal	-10.000 euro + betrouwbaarheidswinst

Hoewel het wegproject en het OV-project hetzelfde effect op de welvaart van Nederland hebben, zal de lezer van het rapport snel geneigd zijn om te concluderen dat het wegproject beter scoort dan het OV-project. Er is sprake van oneerlijke concurrentie. Er is één uitzondering op deze regel. In de MKBA Uithoftram (Ecorys, 2011) zijn betrouwbaarheidswinsten gemonetariseerd meegenomen via de standaardafwijking bij wachttijdverdeling bij haltes. Maar dan nog is er sprake van oneerlijke concurrentie. De bewijslast voor OV-projecten blijft onevenredig hoog ten opzichte van wegprojecten. Bij wegprojecten mag je met een gratis vuistregel van 25% rekenen, terwijl je bij een OV-project een model moet draaien met alle onderzoekslast van dien. Deze onevenwichtigheid wordt alleen maar groter nu een onderzoeksproject is opgestart, waarin men een module ontwikkelt voor betrouwbaarheidseffecten bij wegprojecten in het LMS/NRM. In de nabije toekomst kunnen betrouwbaarheidswinsten voor wegprojecten gemonetariseerd in een MKBA worden meegenomen, nadat zij met een standaardmodel worden geschat. Voor OV-projecten kun je de betrouwbaarheidswinsten kwalitatief meenemen of moet een apart model worden opgetuigd. Dat betrouwbaarheid van OV-projecten nog niet fatsoenlijk in een MKBA kan worden meegenomen is vooral problematisch, vanwege het feit dat de NS wordt gestuurd op betrouwbaarheid. Een project met een positief effect op betrouwbaarheid, maar een negatief effect op reistijd scoort slecht in een MKBA.

Aanbeveling: onderzoek naar het inschatten van het effect van maatregelen in het OV op betrouwbaarheid moet direct worden opgestart. Als hier niet voor wordt gekozen, moet er ook een vuistregel komen voor betrouwbaarheidseffecten van OV-projecten om het evenwicht te herstellen. Omdat het verband tussen reistijdwinst en betrouwbaarheid bij OV-maatregelen relatief casus-specifiek is in vergelijking met wegprojecten, is het in dat geval verstandig om verschillende vuistregels per type OV-maatregel te ontwikkelen.

4.2 Wel VoT voor de pleziervaart maar niet voor de fiets?

Het is zeer opmerkelijk dat de VoT van pleziervaart is ingeschat in deze studie, terwijl de VoT van fietsers niet is ingeschat. Dit, terwijl we in het buitenland voortdurend goede sier maken met onze fietscultuur. De VoT voor wegverkeer proberen we tot achter de komma uit te rekenen, maar we weten niets over de VoT van deze relatief gezonde en milieuvriendelijke vervoerswijze.

Bij MKBAs van het openstaan van bruggen voor pleziervaart, kun je effecten voor de pleziervaart wel in euro's uitdrukken, maar effecten voor fietsers niet. De kengetallen voor de pleziervaart zullen door het ministerie vooral worden gebruikt voor grote sluisprojecten waar er concurrentie is tussen binnenvaart en pleziervaart, maar in de praktijk is het niet onaannemelijk dat er ook MKBAs worden uitgevoerd, waarbij er sprake is van effecten voor pleziervaarders en fietsers. Börjesson en Eliasson (2012) bespreken een eerste aanzet voor het bepalen van de VoT voor fietsers in Zweden.

4.3 Kunnen nationale waarden worden toegepast in alle MKBA's?

Naast het opstarten van een onderzoek naar het inschatten van het effect van maatregelen in het OV op betrouwbaarheid en de Value of Time voor fietsers stel ik voor om een derde onderzoek op te starten. Ik stel namelijk dat het belangrijk is om uit te zoeken voor welke infrastructuurprojecten men de nationale waarden¹¹ zonder verder onderzoek in de MKBA kan toepassen en voor welke infrastructuurprojecten men er goed aan zou doen om (naast de nationale waarden) casus-specifieke VoTs en VoRs in te schatten.

Ik stel namelijk dat het bij bijzondere projecten (projecten waarvoor geldt dat mensen waarop het project effect heeft een stuk lagere of hogere VoT en/of VoR hebben dan gemiddeld) belangrijk is om ook de casus-specifieke VoT en VoR in te schatten. De nationale kengetallen kloppen namelijk niet voor dergelijke projecten. Het gevaar van het toepassen van nationale VoTs en VoRs in MKBAs voor dergelijke projecten is dat je de welvaartseffecten van een project sterk over- of onderschat. Door het keuze-experiment voor een aantal potentiële gebruikers van het project te herhalen kan men de VoT en de VoR – en dus de welvaartseffecten voor het specifieke project – mogelijk substantieel beter inschatten in een MKBA. De kosten van een dergelijk experiment bedragen misschien een paar duizend euro, maar deze kosten verdienen je waarschijnlijk terug als het gaat om een project van meerdere miljoenen euro's. Met name als de baten-kosten ratio rond de 1 ligt en de VoT en de VoR die je kiest bepalend zijn voor de uitkomst van de MKBA (positief of negatief).

Een nadeel is dat deze oplossing beleidstechnisch mogelijk problemen oplevert. De oplossing leidt namelijk al snel tot strategisch gedrag. Je krijgt waarschijnlijk de volgende situatie: de initiatiefnemer van project A laat een casus-specifieke VoT inschatten. Als deze hoger is dan de nationale VoT, neemt hij deze mee in de MKBA. Is de casus-specifieke VoT lager dan de nationale VoT, dan kiest men voor de nationale VoT. In Zweden heeft men een oplossing verzonnen voor dit probleem. Je moet verplicht een MKBA uitvoeren gebaseerd op nationale waarden, daarnaast mag de initiatiefnemer van het project ook gevoeligheidsanalyses doen met casus-specifieke informatie. Uiteindelijk besluit de minister welke informatie hij of zij gelooft. Mijn idee is dat het idealiter niet alleen mogelijk is voor voorstanders om een gevoeligheidsanalyse op basis van casus-specifieke informatie uit te voeren, maar ook voor tegenstanders, als deze bijvoorbeeld het idee hebben dat de VoT lager is dan het nationale gemiddelde.

Mijn voorstel is om onderzoek op te starten waarin twee vragen centraal staan. Ten eerste de vraag: wat zijn voorbeelden van bijzondere projecten waarbij het aannemelijk is dat de VoT en/of VoR (substantieel) hoger of lager ligt dan de nationale waarden? De tweede vraag: hoe kan het beste om worden gegaan met dergelijke bijzondere projecten? Mogelijke opties: 1) Je doet een gevoeligheidsanalyse met een lagere (hogere) VoT en VoR in de MKBA; 2) Je herhaalt het keuze-experiment onder potentiële gebruikers van het project en schat casus-specifieke VoTs en VoRs wanneer het project (bijvoorbeeld) meer dan 10 miljoen euro kost en/of de VoT en de VoR die je kiest bepalend is voor de uitkomst van de MKBA (positief of negatief); 3) de Zweedse oplossing.

In de status quo heeft men in het KiM-rapport (2013, blz. 24) al een regel opgesteld, waarmee men aangeeft welke projecten als 'bijzonder' worden aangemerkt als het gaat om het inschatten van betrouwbaarheidseffecten: *“voor projecten die specifiek tot doel hebben de betrouwbaarheid van reistijden te vergroten, is het dus belangrijk de opslagregel zo veel mogelijk te vermijden. Juist voor dit soort projecten moeten de projecteffecten op de reistijdbetrouwbaarheid nauwkeurig in kaart worden gebracht en worden gewaardeerd met de VoR”*. Deze regel is echter onscherp. Wat is precies nauwkeurig in kaart brengen? Ook kan de regel zeer strategisch worden gebruikt. Als de casus-specifieke VoR hoger is dan de nationale VoR geeft de initiatiefnemer van het project aan dat het doel van het project is om betrouwbaarheid te vergroten en dus de reistijdbetrouwbaarheid nauwkeurig in kaart moet worden gebracht en andersom.

¹¹ In deze sectie ga ik ervan uit dat de nationale waarden niet ter discussie staan. Zie de rest van het paper voor discussie over de nationale waarden.

Het onderzoeksvoorstel dat ik hierboven doe, zorgt ervoor dat deze regel concreter kan worden ingevuld. Voor de VoT geldt dat de rapporten over de VoTVoR-studie niet ingaan op de vraag voor welk type projecten je casus-specifieke informatie zou moeten genereren.

5. Conclusie

Dit paper bediscussieert een aantal claims van de studie 'de maatschappelijke waarde van kortere en betrouwbaardere reistijden'. Er wordt onder andere kritisch gereflecteerd op de claims dat de studie representatief en state-of-the-art is en dat de resultaten in lijn zijn met internationale waarderingsstudies. Samenvattend stelt dit paper dat de VoTVoR-studie een aantal beperkingen heeft en dat de onzekerheden rondom de gevonden VoT en de gevonden VoR best groot zijn. Is dit nu een reden om te concluderen dat de studie niet bruikbaar is en dus in de prullenbak moet? Het antwoord hierop is: 'Absoluut niet!' Het hoofddoel van dit paper is het voeden en scherpstellen van de discussie over het rapport van het KiM en de achtergrondstudie en niet om de suggestie te wekken dat de kwaliteit van de studie onder de maat is, want dit is mijn mening ook niet. Waarom is het dan zo nuttig om de mitsen en maren van deze studie te bespreken?

De verschillende rapporten over de studie 'maatschappelijke waarde van kortere en betrouwbaardere reistijden' besteden voldoende aandacht aan de sterke punten van de VoTVoR-studie. Nu is er ook voldoende informatie over de beperkingen beschikbaar. Alle bouwstenen voor een goede discussie (sterke punten, beperkingen) zijn nu voorhanden. Of dit paper uiteindelijk tot discussie zal leiden en wat de uitkomsten van een dergelijke discussie zouden moeten zijn, blijven uiteraard beleidsvragen.

Waar moet dit paper verder toe leiden? Het paper geeft drie aanbevelingen voor verder onderzoek:

- Onderzoek naar het modelleren van betrouwbaarheid van openbaar vervoer in het LMS/NRM moet direct worden opgestart. Zo kunnen betrouwbaarheidseffecten van OV-projecten in de toekomst gemakkelijk worden meegenomen in de MKBA. Hoewel deze lacune al in 2004 werd geadresseerd (Aanvulling op de Leidraad OEI Directe effecten, 2004) is er anno 2013 nog steeds geen onderzoek opgestart om dit probleem op te lossen;
- Opstarten onderzoeksproject naar de VoT en VoR voor fietsverkeer;
- Opstarten onderzoeksproject gericht op het beantwoorden van twee vragen: 1) Wat zijn voorbeelden van bijzondere projecten waarbij het aannemelijk is dat de VoT en/of VoR (substantieel) hoger of lager ligt dan de nationale waarden? 2) Hoe kan het beste om worden gegaan met dergelijke bijzondere projecten?

Daarbij geeft dit paper aanbevelingen die zich richten op de toepassing van de VoTVoR-studie. De aanbevelingen zijn in twee groepen in te delen: a) beperkingen en onzekerheden verkleinen door meer onderzoek te doen; b) beperkingen en onzekerheden beter communiceren. Hieronder bespreek ik een aantal aandachtspunten bij deze aanbevelingen.

5.1. Beperkingen en onzekerheden verkleinen met meer onderzoek

Dit paper pleit een aantal keer voor een geavanceerder ontwerp van het keuze-experiment en een uitgebreidere dataverzameling. Een keerzijde van dit pleidooi is dat de onderzoekslast omhoog gaat. Is het vergroten van de onderzoeksbudgetten voor dergelijke belangrijke studies een panacee? Dit valt te betwijfelen. Nicolaisen (2012) stipt in zijn promotieonderzoek aan dat onderzoekers geneigd zijn extra onderzoeksbudget vooral te besteden aan het verbeteren van elementen die zij zelf belangrijk vinden. Nicolaisen vroeg aan een aantal econometristen en ingenieurs of hogere onderzoeksbudgetten hen zouden helpen om transparantere rapporten te schrijven over hun onderzoek. De respondenten gaven aan dat zij het als hun taak zien om ervoor te zorgen dat de kwaliteit van de modelschattingen zo hoog mogelijk is. Het budget van een project werd volgens de respondenten in de praktijk vaak volledig opgemaakt aan het verbeteren van de kwaliteit van de modelschattingen. Documentatie en communicatie hadden nauwelijks prioriteit. Een respondent maakte dit

duidelijk met de volgende quote: *'A traditional model builder is so focused on making sure that things fit together. You check everything again and again and re-estimate the model a million times... and this variable here, it needs to be in there. Is it logarithmic or just linear? There's all kinds of that stuff. We'll spend the entire budget, no problem. If I was given triple the budget I would spend that as well and produce just as poor documentation'* (Nicolaisen, 2012. Blz. 197). Uit dit punt in Nicolaisen's thesis valt af te leiden dat een hoger onderzoeksbudget voor de VoTVoR-studie er mogelijk voor had gezorgd dat men op zoek was gegaan naar een nog geavanceerdere manier van modelleren. Aan de beauty-contest van MNL mean dispersion modellen, log WTP modellen en panel-latent class modellen, zouden nog een aantal modellen worden toegevoegd om nog hogere t-waarden, een nog hogere Rho^2 te vinden en nog meer loglikelihood-punten te scoren. Mijn stelling is dat het uitbreiden van het onderzoeksbudget alleen van nut is onder de voorwaarde dat deze uitbreiding zich richt op het versterken van de zwakke plekken in het onderzoek en niet op het nog verder verbeteren van onderdelen in het onderzoek die al zeer goed zijn. Een te grote focus op één kwaliteitsaspect van het onderzoek kan op verschillende manieren worden ondervangen. Bijvoorbeeld door ervoor te zorgen dat de reviewcommissie op de verschillende kwaliteitsaspecten ingaat en door ervoor te zorgen dat in het onderzoeksteam, de reviewcommissie en in het team van opdrachtgevers verschillende soorten onderzoekers vertegenwoordigd zijn. Met name de rol van de opdrachtgever is bepalend. Als deze vraagt om excellente modelschattingen is het waarschijnlijk dat het onderzoeksbudget hier aan wordt besteed.

5.2. Beperkingen en onzekerheden beter communiceren

Dit paper stelt dat de beperkingen en onzekerheden rondom de studie veel beter moeten worden gecommuniceerd. Het gevaar van het niet of beperkt communiceren van de beperkingen en onzekerheden is tweeledig. Ten eerste bestaat het gevaar dat de resultaten van de studie als een heilige graal worden gebruikt. Bij MKBAs voor sluisprojecten in Friesland, een nieuwe Kaagbaan, OV-projecten buiten de Randstad en OV-lijnen waar vooral studenten gebruik van maken worden de gevonden VoT en VoR klakkeloos toegepast, zonder dat men zich afvraagt of dit wel verantwoord is. Ten tweede bestaat het gevaar dat het vertrouwen in de studie afneemt. Sceptische tot gemiddeld sceptische mensen verliezen snel het vertrouwen in studies die schijnzekerheden verkopen en niet eerlijk zijn over hun beperkingen. Het is dus zaak om de beperkingen te communiceren. Als men ervoor kiest om de beperkingen beter te communiceren is het belangrijk om dit op een subtiele manier te doen. Er moet gezocht worden naar een evenwicht tussen 'niet communiceren van beperkingen en onzekerheden' en 'overmatig communiceren van beperkingen en onzekerheden'. Tegelijkertijd is het belangrijk om de implicaties van de onzekerheden en beperkingen te benoemen. Bijvoorbeeld door zinnen op te nemen als: *'doe gevoeligheidsanalyses met een VoT voor wegverkeer van 6 euro en 12 euro'* en *'overweeg het keuze-experiment onder een aantal Friese pleziervaarders te herhalen, indien u een MKBA wilt maken voor een sluisproject in Friesland.'* Met name het verplicht stellen van gevoeligheidsanalyses op de VoT en de VoR in MKBA's is een belangrijke aanbeveling die volgt uit dit paper. Het is opmerkelijk dat in de huidige MKBA-praktijk vaak gevoeligheidsanalyses worden uitgevoerd op de discontovoet en de constructiekosten van een project, maar nauwelijks op de Value of Time.

Daarbij is het belangrijk om 'onzekerheid' ook als mogelijke verklaring op te nemen bij de discussie van de resultaten. Aan deze zin in Warfemmius et al. (2013, blz. 13): *'gemiddeld is de nieuwe reistijdwaardering voor het autoverkeer ruim 16 procent lager dan de huidige. Naast de methodebreuk, kan een plausibele verklaring van de lagere waardering bij het autoverkeer het toenemend gebruik van de mobiele telefoon tijdens de reis zijn'* zou de volgende zin kunnen worden toegevoegd: *'een andere verklaring voor de lagere reistijdwaardering is de onzekerheid rondom de parameters. De standaarddeviatie is immers 2.7'.*

Referenties:

- Axhausen, K.W., Hess, S., König, A., Agay, G., Bates, J. and Bierlaire, M. (2006) State of the Art Estimates of the Swiss Value of Travel Time Savings. Institute of Transport Planning and Systems (IVT), ETH, Zurich.
- Armstrong, J.S., Overton, T., 1977. Estimating nonresponse bias in Mail Surveys. *Journal of Marketing Research* 15. Page 396-402.
- Börjesson, M., Eliasson, J. 2012. *Transportation Research Part A*. 46. Pages 673 – 683.
- Brömmenstroet, M. C. G., te., Brugge, R. ter., Bertolini, L., 2012. 40 jaar spookrijden in de vervoersplanologie: iatrogenese, naïeve interventies en een actieve rol voor het CVS. Bijdrage aan het Colloquium Vervoersplanologisch Speurwerk. 2013, Rotterdam.
- CBS, 2010. Onderzoek Verplaatsingen in Nederland 2010.
- CPB, 2004. Economische toets op de Nota Mobiliteit, Document nr. 65, Den Haag.
- Ecorys, 2011. MKBA Uithoflijn, rapportage uitkomsten. Opdrachtgever: projectorganisatie Uithoflijn. Rotterdam 9 mei 2011.
- Ferreira, A., Beukers, E., Brömmenstroet, M. C. G., te. 2012. Accessibility is gold, mobility is not: a proposal for the improvement of Dutch transport-related cost-benefit analysis. *Environment and Planning B: Planning and Design* 2012, volume 39, pages 683-697.
- Fosgerau, M., Hjorth, K. and Lyk-Jensen, S.V. (2007) The Danish Value of Time Study: Final Report. Danish Transport Research Institute, Knuth-Winterfeldt Allé, Bygning 116 Vest, 2800 Kgs. Lyngby.
- Hague Consulting Group, 1998. The second Netherlands' value of time study: final report, in opdracht van de Adviesdienst Verkeer en Vervoer, Den Haag.
- Hjorth, K., Fosgerau, M., 2012. Using prospect theory to investigate the low marginal value of travel time for small time changes. *Transportation Research Part B* 46. P. 917-932.
- Huys, M. (2011) Building Castles in the (Dutch) Air: Understanding the Policy Deadlock of Amsterdam Airport Schiphol 1989 – 2009. Ph.D. Dissertation, TU Delft.
- Kennisinstituut voor Mobiliteitsbeleid, 2013. De maatschappelijke waarde van kortere en betrouwbaardere reistijden, Den Haag.
- Levine, L., 2006. A Geography of time. The temporal misadventures of a social psychologist, or how every culture keeps time just a little bit differently. Published by Oneworld Publications.
- Mackie, P., Preston, J., 1998. Twenty-one sources of error and bias in transport project appraisal. *Transport Policy* 5 (1), 1-7.
- Mackie, P.J., M. Wardman, A.S. Fowkes, G. Whelan, J. Nellthorp and J. Bates, 2003. Values of travel time savings in the UK, ITS Leeds and John Bates Services report for the UK Department for Transport, Leeds.
- Ministerie van Verkeer en Waterstaat, 2004. Directe Effecten Infrastructuurprojecten: Aanvulling op de Leidraad OEI, Den Haag.
- Mishan, E.J., 1988. *Cost-Benefit Analysis*, Allen and Unwin, London.
- Mouter, N., 2012. Voordelen en nadelen van de Maatschappelijke kosten- en batenanalyse nader uitgewerkt. Bijdrage aan het Colloquium Vervoersplanologisch Speurwerk. November 2012;
- Mouter, N., Annema, J.A., Van Wee, B., 2013a. Ranking the substantive problems in the Dutch Cost-Benefit Analysis practice. *Transportation Research Part A* 49, 241-255.
- Mouter, N., forthcoming MKBA international, lessen uit een vergelijking van de Nederlandse MKBA-praktijk met vier andere MKBA-praktijken. In opdracht van het Planbureau voor de Leefomgeving.
- Ramjerdi, F., Flügel, S., Samstad, H. and Killi, M., 2010 Value of time, safety and environment in passenger transport – Time. TØI report 1053B/2010, Institute of Transport Economics, Oslo
- Significance, VU University Amsterdam, John Bates Services, TNO, NEA, TNS NIPO & PanelClix, 2012. Values of time and reliability in passenger and freight transport in The Netherlands, A report for the Ministry of Infrastructure and the Environment, Den Haag.
- Tseng, Y.Y., E.T. Verhoef, G.,C. de Jong, M. Kouwenhoven and A.I.J.M. van der Hoorn, 2009. A pilot study into the perception of unreliability of travel times using in-depth interviews, *Journal of Choice Modelling*, 2(1), 8-28.
- Wardman, M., P. Chintakayala, G.C. de Jong and D. Ferrer, 2012. European wide meta-analysis of values of travel time; Paper prepared for EIB, ITS Leeds.
- Warfemmius, P., Kouwenhoven, M., de Jong, G.C., 2013. De maatschappelijke waarde van kortere en betrouwbaardere reistijden. Bijdrage aan het Colloquium Vervoersplanologisch Speurwerk. 2013, Rotterdam.