

5. *Wie het weet die moet dat zeggen* *Het aggregeren van informantgegevens in organisationeel marktonderzoek^{1,2}*

G.H. VAN BRUGGEN, G.L. LILIEN en M. KACKER

SAMENVATTING

In marktonderzoek worden gegevens over organisaties vaak verzameld door middel van het ondervragen van zogenaamde informanten. In dit artikel betogen we dat het gebruik van meerdere informanten per organisatie in plaats van één enkele informant de kwaliteit van responsgegevens en daarmee de kwaliteit van onderzoek zal verhogen. Informanten kunnen echter van mening verschillen. Dit leidt tot de vraag hoe opinies geaggregeerd moeten worden. We beschrijven een aantal aggregatiemethoden die gebruik maken van de inschattingen van informanten van de nauwkeurigheid van hun respons en van informatie over de inschattingen van informanten relatief ten opzichte van andere informanten. Onze methoden zijn effectief en relatief gemakkelijk toepasbaar in de praktijk van organisationeel marktonderzoek.

1. INLEIDING

In veel marketing- en marktonderzoekstudies verzamelen onderzoekers gegevens over verschillende variabelen (zoals marketinginspanningen, verkopen en marktaandeel) om vervolgens na te gaan of er een verband tussen deze variabelen bestaat. In marktonderzoek onder organisaties, of organisationeel marktonderzoek, is de organisatie of een organisatieonderdeel het object van onderzoek. Als de juiste gegevens verkregen kunnen worden door gebruik te maken van archieven, bedrijfsadministraties, jaarverslagen etc. dan zullen onderzoekers dit doen. Vaak zullen gegevens echter niet gemakkelijk te verzamelen zijn of zijn bronnen binnen organisaties moeilijk toegankelijk (Kumar, Stern en Anderson 1993). Onderzoeksgegevens zullen dan verzameld moeten worden door gebruik te maken van informanten. Deze informanten geven hun inschatting over de waarde van organisatievariabelen. Informanten rapporteren dus hun eigen subjectieve, retrospectieve inschattingen met betrekking tot eigenschappen van de organi-

¹ Deze bijdrage is gebaseerd op het artikel "Informants in Organizational Marketing Research: Why Use Multiple Informants and How to Aggregate Responses" van Gerrit H. van Bruggen, Gary L. Lilien en Manish Kacker dat in November 2002 in het *Journal of Marketing Research* gepubliceerd is. Voor meer achtergronden van de in deze bijdrage beschreven methoden en een uitgebreide beschrijving van de empirische studie verwijzen wij naar dit artikel.

² De auteurs danken redacteur Ale Smidts voor zijn waardevolle commentaar op een eerdere versie van dit artikel.

satie (bijvoorbeeld verkopen en winstcijfers). Zij verschillen hierin van respondenten die informatie over zichzelf als individu verschaffen (bijvoorbeeld hun persoonlijke beoordeling van de kwaliteit van de organisatie).

Onderzoekers die informatie over organisationele variabelen verzamelen door gebruik te maken van informanten worden geconfronteerd met twee belangrijke vragen: 1) moet gebruik gemaakt worden van *één* of van *meer* (multiple) informanten; en indien meer dan één informant gebruikt wordt 2) hoe moeten de gegevens van meerdere informanten dan geaggregeerd worden?

Het is vaak gemakkelijk en bovendien goedkoop om per organisatie slechts één informant om informatie te vragen. Dit gebeurt dan ook vaak in de praktijk. Diverse studies laten echter zien dat het gebruik van meerdere informanten de kwaliteit van de verzamelde gegevens zal vergroten (Seidler 1974; Hogarth 1978; Hill 1982). Diverse onderzoekers bepleiten daarom het gebruik van meerdere informanten in organisationeel marktonderzoek. In deze bijdrage willen we ingaan op de vraag *waarom* het gebruik van gegevens van meerdere informanten de nauwkeurigheid van informatie en van gevonden relaties tussen variabelen vergroot en *hoe* bij gebruik van meerdere informanten de respons van deze informanten geaggregeerd moet worden.

2. HET GEBRUIK VAN INFORMANTEN: ÉÉN OF MEERDERE?

Objectieve organisationele variabelen zoals de verkopen en winst van een bedrijf worden gemeten op het niveau van de organisatie (of business unit). Deze variabelen hebben als kenmerk dat een juiste waarde (true score) voor de organisatie bestaat. De door informanten opgegeven waarde van variabelen zal echter kunnen afwijken van de juiste waarde doordat er sprake is van meetfouten (bijvoorbeeld omdat informanten zich bepaalde informatie niet goed meer herinneren). Deze meetfouten bevatten vaak een systematische - en een random component. Samenvattend:

$$\text{Respons van Informant} = \text{Juiste Waarde} + \text{Meetfout}$$

$$\text{Meetfout} = \text{Systematische Meetfout} + \text{Random Meetfout}$$

De random meetfout ontstaat omdat van informanten gevraagd wordt om als zogenaamde "key informant" inschattingen te maken over de waarde van organisatievariabelen. Het maken van een accurate inschatting van deze waarden is vaak een complexe zaak (Philips 1981) en informanten kunnen hierbij fouten maken. Random meetfouten kennen geen systematisch patroon. De verwachte waarde van de random meetfout is dan ook nul.

De systematische meetfout in de respons van informanten kan zowel individuele oorzaken als organisationele oorzaken hebben. Individuen kunnen systematisch een te hoge of te lage response voor variabelen geven, bijvoorbeeld vanwege sociale wenselijkheid. Een organisationele oorzaak die tot systematische meetfouten kan leiden kan bijvoorbeeld liggen in de afdeling in de organisatie waar de informant werkzaam is. Zo kun-

nen medewerkers van de marketingafdeling de marktgerichtheid van een organisatie systematisch hoger inschatten dan medewerkers van andere afdelingen.

De omvang van de foutencomponent kan substantieel zijn. Zo stelt Philips (1981) dat de informatie van informanten over de waarde van variabelen vaak een variantie kent die voor meer dan 50% is toe te schrijven aan systematische - en random meetfouten.

Onderzoekers zijn vaak geïnteresseerd in relaties tussen variabelen. De relaties die zij vinden zijn die tussen de gerapporteerde responsgegevens van de informanten. Omdat deze respons een foutencomponent bevat zullen de gerapporteerde correlaties niet alleen afhankelijk zijn van de correlaties tussen de werkelijke, juiste waarden van de variabelen, maar ook van de correlaties tussen de foutencomponenten en de omvang van deze foutencomponenten. Om er voor te zorgen dat de gevonden relatie tussen constructen zo goed mogelijk weergeeft wat de relatie tussen de werkelijke waarden van deze variabelen is, is het belangrijk dat de omvang van de foutencomponenten en de correlatie tussen deze foutencomponenten zo klein mogelijk is.

Het gebruik van meerdere informanten en het simpelweg middelen van de respons van deze verschillende informanten tot één enkele organisatiescore leidt tot een verkleining van de random meetfout in organisationele variabelen. De random meetfouten in de respons van individuele informanten worden zo tegen elkaar weggemiddeld. Hoe groter het aantal informanten, hoe kleiner de verwachte random meetfout zal worden. Het *optimale* aantal informanten hangt af van de kosten van het verkrijgen van additionele informanten ten opzichte van de kosten van de overblijvende meetfout. Onderzoek van Ashton en Ashton (1985) op het gebied van voorspellingen laat zien dat een aantal informanten dat tussen twee en vijf ligt effectief is. Libby en Blashfield (1978) stellen dat door het gebruik van drie informanten de meeste winst in nauwkeurigheid kan worden gerealiseerd.

Naast een reductie van de random meetfout is het ook belangrijk om de invloed van de systematische meetfout te minimaliseren. Ook hier geldt dat het gebruik van meerdere informanten een effectief middel is. In tegenstelling tot verwachte waarde van de random meetfout zal de verwachte waarde van de systematische meetfout in de respons van individuele informanten niet gelijk aan nul zijn. Bovendien zal de systematische fout voor de verschillende variabelen, waarvoor een informant informatie verschaft, vergelijkbaar zijn. Een informant die systematisch overschat zal voor alle variabelen een te hoge respons geven. Dit leidt tot een positieve correlatie tussen de foutencomponenten in deze verschillende variabelen. De verwachte waarde van de correlatie tussen systematische meetfouten neemt dus toe als responsgegevens voor variabelen van dezelfde informant afkomstig zijn. Dit zal de gevonden correlatie tussen de variabelen die bestudeerd worden beïnvloeden.

Een simpele manier om de correlatie tussen systematische fouten te reduceren is het gebruiken van verschillende informanten per organisatie voor verschillende variabelen. Zo kan bijvoorbeeld een eerste informant in een organisatie gevraagd worden om een inschatting van de marketinginspanningen terwijl een tweede informant gegevens over marketingresultaten verschaft. De respons van verschillende informanten moet hierbij zo onafhankelijk mogelijk van elkaar zijn. Door het gebruik van meerdere onafhankelijke informanten zal de verwachte waarde van de correlatie tussen systematische foutencomponenten afnemen, waardoor de gevonden correlatie dichterbij de werkelijke of

“true” correlatie ligt. Het gebruik van meerdere informanten maakt het bovendien mogelijk om per construct de meest deskundige informant te selecteren. Ook dit zal mogelijke meetfouten verkleinen. Het is immers voor informanten lastig om over alle variabelen de meest betrouwbare informatie te verstrekken.

Samenvattend stellen we dat in organisationeel marktonderzoek het gebruik van meerdere informanten te prefereren is boven het gebruik van één enkele informant per organisatie. Verschillende informanten kunnen over verschillende variabelen rapporteren. Ze kunnen echter ook over dezelfde variabelen rapporteren. Hierbij zullen ze echter vaak verschillende opinies hebben. Dit leidt tot de vraag hoe met de respons van de verschillende informanten moet worden omgegaan om tot één enkele organisatie waarde voor variabelen te komen.

3. HET AGGREGEREN VAN INFORMANTENRESPONS

Verschillende informanten die afkomstig zijn uit één en dezelfde organisatie zullen vaak verschillende meningen hebben. Informanten zullen vaak ook verschillen in de nauwkeurigheid van hun respons. Het is daarom belangrijk om na te denken over de vraag hoe de verschillende opinies geaggregeerd moeten worden. Aggregatie is nodig om één enkele organisatie waarde te verkrijgen. In de literatuur worden zowel behavioral – als mathematische aggregatiemethoden beschreven.

Volgens de behavioral aggregatiemethode zullen de verschillende informanten die over dezelfde organisatie of over hetzelfde organisatieonderdeel rapporteren, discussiëren over de waarde van een variabele waarover zij van mening verschillen. Zij proberen het verschil van mening weg te werken en het zo eens te worden over één groeps waarde (Ferrel 1985). De behavioral aggregatiemethode biedt zo een directe oplossing voor het aggregatieprobleem. De methode vraagt echter substantiële inspanningen en coördinatie van de informanten. Verder maakt ze het lastig voor informanten om vertrouwelijk en anoniem informatie te geven. Tevens kan de resulterende groeps waarde mogelijk beïnvloed worden door groepsprocessen en machtsverhoudingen. Tenslotte kan het zijn dat de in het overleg meest invloedrijke groepsleden niet de meest accurate zijn (Lar-réché en Moinpour 1983). Al deze factoren kunnen de praktische toepasbaarheid en de nauwkeurigheid van de behavioral aggregatiemethode negatief beïnvloeden.

Omdat de van de informanten gevraagde inspanningen de behavioral aggregatiemethode in de praktijk moeilijk uitvoerbaar maken ontwikkelden Kumar, Stern en Anderson (1993) de zogenaamde hybride consensus-aggregatiemethode. Volgens deze methode wordt voor de variabelen waarover geen grote meningsverschillen tussen de informanten bestaan simpelweg het rekenkundig gemiddelde berekend. Voor variabelen waarvoor grotere verschillen bestaan zullen de leden van een groep door middel van discussie overeenstemming proberen te realiseren. Empirisch onderzoek van Kumar, Stern en Anderson (1993) laat echter zien dat het resultaat van discussies tussen groepsleden vaak dichter ligt bij de mening van informanten met een hiërarchisch gezien hogere positie. Een alternatieve aggregatiemethode die minder last heeft van groepsprocessen is de Delphi methode. Deze methode is echter kostbaar vanwege de hoeveelheid tijd en inspanningen die zij vraagt van de informanten.

Mathematische aggregatiemethoden bieden een alternatief voor de behavioral methoden. Een veelgebruikte mathematische aggregatiemethode is het simpelweg berekenen van het rekenkundig gemiddelde over de verschillende meningen van verschillende informanten. Als informanten echter substantieel van mening verschillen dan zal deze methode zelden tot de meest nauwkeurige groepswaarde leiden (James 1982). Hill (1982) laat zien dat de gemiddelde groepswaarde vaak wel nauwkeuriger is dan de respons van individuele groepsleden maar dat door meer geavanceerde statistische aggregatiemethoden vaak nog nauwkeuriger resultaten verkregen kunnen worden. Zogenaamde Bayesiaanse methoden (Morris 1977) integreren de opinies van verschillende informanten op elegante en theoretische geavanceerde wijze. Ze zijn echter niet altijd even gemakkelijk toepasbaar in praktijksituaties omdat ze veel inspanningen vragen van zowel informanten, die gecompliceerde kansverdelingen moeten inschatten (Clemen en Winkler 1993), als van onderzoekers.

De inspanningen die bij zowel de behavioral als bij de Bayesiaanse aggregatiemethoden van informanten gevraagd wordt hebben ertoe geleid dat ze in de praktijk slechts beperkt worden toegepast. Het berekenen van het rekenkundig gemiddelde is eenvoudig maar zal alleen effectief zijn als er geen systematische foutencomponent in de respons van informanten aanwezig is. Hiervan zal in organisationeel onderzoek zelden sprake zijn.

Als er sprake is van systematische meetfouten in de respons van individuele informanten, dan is het belangrijk om vast te stellen wat de oorzaak van deze fouten is en is het nodig per organisatie(onderdeel) die informant te identificeren waarvan de systematische responsfout het kleinst is. Als met zekerheid vastgesteld kan worden welke informant de kleinste systematische fout in zijn/haar respons kent, dan moet de respons van die informant gebruikt worden als groepswaarde. Als de nauwkeurigheid van de groepsleden niet met zekerheid kan worden vastgesteld, dan moet een gewogen gemiddelde bepaald worden, waarbij de informanten waarvan verwacht wordt dat die meer nauwkeurige informatie verschaffen een groter gewicht krijgen dan de informanten die mogelijk minder nauwkeurig zijn.

Het is nu de vraag hoe we vast kunnen stellen wat de nauwkeurigheid van een individuele informant is en hoe we deze informatie vervolgens kunnen gebruiken in een aggregatieprocedure.

4. METHODEN VOOR DE AGGREGATIE VAN INFORMANTEN-RESPONS

In deze paragraaf introduceren we drie methoden voor het aggregeren van gegevens van informanten. Dit zijn: (1) het ongewogen rekenkundig groepsgemiddelde (dit gemiddelde geldt als referentiewaarde); (2) een gewogen groepsgemiddelde waarbij de gewichten worden afgeleid uit de door de informanten verstrekte gegevens; en (3) een gewogen groepsgemiddelde waarbij de gewichten worden afgeleid uit door de informanten verstrekte informatie met betrekking tot hun eigen vertrouwen in de door hen verstrekte gegevens.

1. Het ongewogen groepsgemiddelde (UNWMEAN)

Volgens onze eerste methode berekenen we een groepswaarde voor een variabele door

het rekenkundig gemiddelde van de individuele scores van groepsleden te bepalen. De waarde van het ongewogen groeps­gemiddelde voor organisatie­variabele X (in de empirische toepassing die later zal worden besproken analyseren we acht organisatie­variabe­len, dus $X = 1, \dots, 8$), $UNWMEAN_{xi}$ van groep i, wordt als volgt berekend:

$$UNWMEAN_{xi} = \frac{\left(\sum_{j=1}^{n_i} X_{ij} \right)}{n_i} \quad (1)$$

Hierbij geldt

X_{ij} = de waarde van variabele X zoals gegeven door informant j in groep i

n_i = aantal informanten in groep i

Stel dat drie informanten rapporteren over het marktaandeel van een bedrijf. Informant 1 schat dit aandeel in op 10%, informant 2 op 20% en informant 3 op 21%. Het ongewogen groeps­gemiddelde voor dit bedrijf is dan 17% $((10+20+21)/3)$.

2. Het met responsgegevens van informanten gewogen groeps­gemiddelde (WDMEAN)

De tweede methode voor het aggregeren van gegevens van informanten gaat er vanuit dat de *mate van overeenstemming* tussen informanten uit een groep kan worden gebruikt bij het bepalen van een groeps­waarde. De afwijking van een individueel groeps­lid van de opinie van de overige groeps­leden zegt mogelijk iets over de nauwkeurigheid van dit lid. Stel, bijvoorbeeld, dat twee informanten uit een groep van drie leden dezelfde inschattingen voor de waarde van een variabele geven, terwijl een derde groeps­lid een substantieel afwijkende waarde geeft. In een dergelijke situatie wordt de informatie van de twee informanten die het met elkaar eens zijn zwaarder gewogen dan de informatie van de informant die een afwijkende mening heeft. Deze aanpak gaat er dus vanuit dat de juiste waarde in het algemeen dichter bij de inschattingen van de meerderheid in de groep ligt dan bij de inschatting van groeps­leden met een afwijkende mening.

Voor de berekening van $WDMEAN_{xi}$ berekenen we eerst de gewichten voor ieder van de informanten binnen de groep. Dit doen we door eerst $DIST_{xij}$ te berekenen. Deze variabele geeft de absolute afstand tussen informant j's inschatting van de waarde van variabele X en het ongewogen, rekenkundig gemiddelde van groep i waartoe informant j behoort. $DIST_{xij}$ wordt als volgt berekend:

$$DIST_{xij} = |X_{ij} - UNWMEAN_{xi}| \quad (2)$$

Voor het eerder geïntroduceerde getallen­voorbeeld zal de waarde van DIST voor de eerste informant 7 (=|10-17|) bedragen. Voor informant 2 is deze waarde 3 (=|20-17|) en voor de derde informant is deze 4 (=|21-17|).

Het gewicht ($WEIGHT_{xij}$) dat wordt toegekend aan de inschatting van informant j van de waarde van variabele X is omgekeerd evenredig aan de absolute afstand van deze inschatting tot het ongewogen groeps­gemiddelde voor groep i, ten opzichte van deze afstand voor de andere informanten uit groep i. We berekenen het gewicht van informant j's inschatting voor variabele X ($WEIGHT_{xij}$) als volgt:

$$\text{WEIGHT}_{xij} = \left(\frac{\left(\sum_{j=1}^{n_i} \text{DIST}_{xij} \right)^{\alpha}}{\text{DIST}_{xij}} \right) \quad (3)$$

In formule (3) introduceren we parameter α die de defaultwaarde 1 kent. Bij deze waarde van 1 is het gewicht van een informant omgekeerd evenredig aan zijn/haar relatieve afstand tot het ongewogen groepsgemiddelde. De waarde van α kan worden verhoogd wanneer we de gewichten van informanten met een inschatting die dichter bij het ongewogen groepsgemiddelde ligt, ten opzichte van informanten met een afwijkende mening, willen verhogen. Als α de waarde 0 heeft dan wegen de inschattingen van de verschillende groepsleden even zwaar (dit is dus is gelijk aan het ongewogen groepsgemiddelde).

Uitgaande van de defaultwaarde 1 voor parameter α wordt de waarde van WEIGHT voor de eerste respondent uit het getallenvoorbeeld 2 $(=(7+3+4)/7)$. Voor informant 2 wordt deze waarde 4.67 $(=(7+3+4)/3)$ en voor de derde informant 3.5 $(=(7+3+4)/4)$.

We berekenen nu het op responsgegevens van informanten gebaseerde gewogen groepsgemiddelde (WDMEAN_{xi}) voor variabele X voor groep i. De inschatting van ieder individueel groepslid wordt gewogen op basis van zijn relatieve afstand tot het ongewogen groepsgemiddelde. WDMEAN_{xi} wordt als volgt berekend:

$$\text{WDMEAN}_{xi} = \sum_{j=1}^{n_i} \left(\frac{\text{WEIGHT}_{xij}}{\left(\sum_{j=1}^{n_i} \text{WEIGHT}_{xij} \right)} * X_{ij} \right) \quad (4)$$

Volgens deze aggregatiemethode wordt het gewogen groepsgemiddelde berekend als:
 $(2/(2+4.67+3.5))*10 + (4.67/(2+4.67+3.5))*20 + (3.5/(2+4.67+3.5))*21 = 18.38$.

Te zien is dat het op deze wijze berekende groepsgemiddelde dichter bij de waarde van de informanten 2 en 3 (die het met elkaar eens waren) ligt dan het ongewogen groepsgemiddelde. De waarde ligt verder af van informant 1 (die een afwijkende mening had).

3. Het met vertrouwen van informanten gewogen groepsgemiddelde (WCMEAN)

De derde aggregatiemethode gebruikt het vertrouwen van informanten in de nauwkeurigheid van hun inschatting van de waarde van variabele X. De inschattingen van informanten met meer vertrouwen worden hierbij zwaarder gewogen dan de inschattingen van informanten die minder zeker zijn. WCMEAN_{xi} , de waarde voor variabele X voor groep i waar informant j's inschatting wordt gewogen met zijn/haar vertrouwen CONF_{xij} in de nauwkeurigheid van deze inschatting, wordt berekend als:

$$WCMEAN_{xi} = \sum_{j=1}^{n_i} \left(\frac{CONF_{xij}^{\alpha}}{\left(\sum_{j=1}^{n_i} CONF_{xij}^{\alpha} \right)} * X_{ij} \right) \quad (5)$$

In formule (5) maken we opnieuw gebruik van parameter α die het mogelijk maakt om de gewichten die worden toegekend aan informanten met minder of meer vertrouwen in hun inschattingen te manipuleren. Parameter α heeft ook hier de defaultwaarde 1. Als deze waarde verhoogd wordt dan wordt het gewicht van informanten met relatief veel vertrouwen in de eigen inschattingen groter. Als de waarde van α gelijk aan 0 wordt gesteld dan leidt dit tot de berekening van het ongewogen groepsgemiddelde.

Ook deze derde aggregatiemethode kunnen we illustreren aan de hand van het eerder gebruikte getallenvoorbeeld. Stel dat informant 1 op een 9-puntsschaal een vertrouwensscore 7 had (in de inschatting 10), informant 2 een score van 8 en informant 3 een score 6. Het met vertrouwen gewogen groepsgemiddelde is dan: $(7/(7+8+6))*10 + (8/(7+8+6))*20 + (6/(7+8+6))*21 = 16.95$.

Na de beschrijving van de drie aggregatiemethoden bespreken we nu een empirische studie waarin de kwaliteit van de aggregatiemethoden wordt geanalyseerd.

5. HET AGGREGEREN VAN “RECALL” GEGEVENS

5.1. Opzet van de studie

Om de kwaliteit van onze aggregatiemethoden te bepalen was het noodzakelijk om gegevens van informanten te verzamelen in een realistische organisatieomgeving. Hierbij moest het mogelijk zijn om de inschattingen van de informanten te vergelijken met een objectieve, correcte (juiste) waarde van de variabele. De MARKSTRAT simulatie (Larréché en Gatignon 1990) is een omgeving die deze mogelijkheden biedt. MARKSTRAT is in het verleden vaak gebruikt voor het bestuderen van besluitvorming in marketing (zie bijvoorbeeld Glazer, Steckel en Winer 1992). In de MARKSTRAT simulatie dienen deelnemers in groepsverband gedurende een aantal tijdsperiodes beslissingen te nemen over zowel strategische - als tactische marketingvraagstukken voor verschillende elkaar beconcurrerende ondernemingen.

De informanten in onze eerste empirische studie waren 67 marketing studenten die in het kader van hun opleiding deelnamen aan een marketingstrategie cursus aan een universiteit in de Verenigde Staten. De studenten vormden twee, drie of vier-persoons groepen die beslissingen moesten nemen voor één van de vijf bedrijven die actief waren in vier MARKSTRAT bedrijfstakken. In totaal opereerden de 67 informanten in 20 groepen. De groepen namen iedere periode beslissingen na bestudering van de resultaten van voorgaande periodes en van marktonderzoeksrapporten. Alle groepen met deelnemers ontvingen dezelfde informatie en hadden dezelfde hoeveelheid tijd om beslissingen te nemen. De beslissingen werden gelijktijdig door alle groepen genomen. Nadat de groepen een aantal ronden aan de MARKSTRAT simulatie hadden deelgeno-

men werd ieder van de deelnemers gevraagd om *individueel* een vragenlijst in te vullen. Onder deze vragen bevonden zich een achttal vragen die de informanten vroegen om de waarden van MARKSTRAT variabelen te reproduceren (recall data). Het ging hier om de waarden van marketing mix variabelen zoals reclame-uitgaven, prijzen en aantallen vertegenwoordigers waarvoor de groepen beslissingen hadden genomen alsook voor het marketingbudget dat de groepen beschikbaar hadden voor toekomstig beleid. Hiernaast werd de informanten gevraagd om per variabele hun *vertrouwen* in hun eigen inschatting te geven op een 9-puntsschaal. Hier betekende een 1 dat de informanten “absoluut niet zeker” waren van hun inschatting terwijl een 9 een “absolute zekerheid” in de nauwkeurigheid van de inschatting betekende. Om de betrokkenheid van de informanten bij de studie te waarborgen, om hun nauwkeurigheid te stimuleren en om samenwerking tussen de leden van een groep te ontmoedigen, ontvingen de informanten die de meest nauwkeurige inschattingen wisten te geven een beloning. Tenslotte is het belangrijk om twee kenmerken van onze onderzoeksopzet te noemen. Ten eerste waren alle deelnemers studenten die geen hiërarchisch of functioneel verschillende posities in hun MARKSTRAT groep innamen. De informanten hadden dus hetzelfde (homogene) perspectief, wat een hiërarchische of functionele bias onwaarschijnlijk maakt. Ten tweede produceert het MARKSTRAT computerprogramma de werkelijke (juiste) waarden voor alle variabelen waar ook de informanten gegevens over verstrekken. Dit maakte het mogelijk om objectief de nauwkeurigheid van de inschattingen van de informanten vast te stellen.

5.2 Resultaten

De aggregatiemethoden, zoals eerder besproken, werden toegepast op de gegevens zoals de informanten in de verschillende MARKSTRAT groepen produceerden. Op deze manier werd voor ieder van de twintig MARKSTRAT groepen een waarde voor de acht variabelen bepaald. Om de nauwkeurigheid van de verschillende methoden te bepalen, berekenden we per methode hoeveel procent de geaggregeerde waarde gemiddeld van de werkelijke waarde afweek (*Mean Absolute Percentage Error (MAPE)*). De MAPE voor groep i voor een specifieke aggregatiemethode (gemiddeld over de acht MARKSTRAT variabelen) wordt als volgt berekend:

$$MAPE_i = \left(\frac{\sum_{k=1}^8 \left| \frac{\text{Estimated Value of } X_{ki} - \text{Real Value of } X_{ki}}{\text{Real Value of } X_{ki}} \right| * 100\%}{8} \right) \quad (6)$$

In Tabel 1 worden de resultaten voor de verschillende aggregatiemethoden weergegeven.

Tabel 1. Nauwkeurigheid van de drie aggregatiemethoden (standaard afwijking tussen haakjes).

Aggregatie Methode	MAPE
1. ongewogen groepsgemiddelde	16.30% (10.08)
2. met informanten respons gewogen groepsgemiddelde	14.20% (10.22)
3. met informanten vertrouwen gewogen groepsgemiddelde	12.81% (8.37)

De resultaten in Tabel 1 laten zien dat aggregatiemethoden die gebruik maken van gewichten tot een nauwkeuriger bepaling van een groepswaarde leiden ($F=15.45$, $p=.001$). De waarden van MAPE zijn voor methode 2 en 3 kleiner dan voor methode 1. Verder is te zien dat het gebruik van gewichten gebaseerd op het vertrouwen van individuele informanten in hun eigen inschattingen tot nauwkeuriger resultaten leidt dan het gebruik van op de respons gebaseerde gewichten. Het gebruik van op vertrouwen gebaseerde gewichten leidt tot een ruim 20% hogere nauwkeurigheid van de resultaten ($F=20.88$, $p<.001$) vergeleken met het ongewogen groepsgemiddelde (12.81% vs. 16.30%).

De resultaten in Tabel 1 zijn gebaseerd op het gebruik van een standaardwaarde 1 voor parameter α in formules (3) en (5). Het is nu de vraag of de resultaten van de aggregatiemethoden verbeteren als de waarde van α verandert. Voor de aggregatiemethode 2 (op informanten respons gebaseerd) berekenen we daarom de waarde van α in formule (3) die de waarde van MAPE (Formule 6) minimaliseert. Hiervoor maken we gebruik van de Solver module uit Microsoft Excel. We berekenen een optimale waarde voor α voor ieder van de acht MARKSTRAT variabelen (we noemen dit de variabele-specifieke α). Hiernaast berekenen we ook een optimale α die een gelijke waarde kent voor alle acht MARKSTRAT variabelen (de uniforme α). In Tabel 2 presenteren we de resultaten van deze aanpak.

Tabel 2. Nauwkeurigheid (MAPE) van de aggregatiemethoden bij het gebruik van verschillende α waarden (standaardafwijkingen tussen haakjes).

Aggregatie Methode	$\alpha = 1$ default waarde)	uniforme α (geoptimaliseerd over de acht MARKSTRAT variabelen)	variabele specifieke α (geoptimaliseerd per MARKSTRAT variabele)
met informant respons gewogen groepswaarde	14.20% (10.22)	12.45% (10.76) ($\alpha = 25.70$)	12.35% (10.71) (α range = 2.88–77.00)
met informant vertrouwen gewogen groepswaarde	12.81% (8.37)	7.90% (6.24) ($\alpha = 12.87$)	7.53% (6.07) (α range = 5.46–25.80)

De resultaten in Tabel 2 laten zien dat de nauwkeurigheid van de aggregatiemethode die gebruik maakt van de op informanten respons gebaseerde weging met ongeveer 15% (12.45 vs. 14.20) verbeterd kan worden door optimalisatie van de α waarde ($F=9.11$, $p=.007$). De optimale uniforme α waarde is 25.70, terwijl de optimale waarde voor de variabele specifieke α waarde varieert tussen 2.88 en 77.00. Het verschil tussen de MAPE bij het gebruik van variabele specifieke α waarden en die bij een uniforme α waarde is klein (12.35 vs. 12.45). Kennelijk komt de meeste winst in nauwkeurigheid uit het *vergroten van het gewicht van de informanten die overeenstemming vertonen*. Dit gebeurt door een waarde van α te gebruiken die substantieel boven 1 ligt. Hierbij is de nauwkeurigheid van de aggregatiemethode minder gevoelig voor de feitelijke waarde van de α .

De α waarden voor de op informant vertrouwen gebaseerde methoden kunnen op dezelfde manier geoptimaliseerd worden. We berekenen dan een waarde van α in Formule 5 die de waarde van MAPE in Formule 6 minimaliseert. Hierbij bepalen we zowel een optimale uniforme α waarde als een variabele specifieke optimale α waarde. Ook hier laten de resultaten zien dat het vergroten van de gewichten van de informanten met vertrouwen in hun eigen inschattingen tot een substantiële verbetering van de nauwkeurigheid van de aggregatiemethoden leidt. Als we α de waarde 1 geven dan bedraagt de MAPE 12.81. De berekening van een uniforme optimale α waarde leidt tot een daling van de MAPE tot 7.90 terwijl het gebruik van variabele specifieke α waarden leidt tot een MAPE van 7.53. Dit betekent dat door gewogen aggregatie een verbetering in de nauwkeurigheid van de groepswaarden van meer dan 50% verkregen kan worden (7.90 vs. 16.30). Ook hier geldt dat het berekenen van optimale waarden van α per variabele tot nauwelijks betere resultaten leidt dan het gebruik van een uniforme optimale α waarde (7.53 vs. 7.90).

Om onze methoden nu succesvol toe te kunnen passen is het uiteraard noodzakelijk om de waarde van de optimale α te weten. Deze kan bepaald worden door informanten te verzoeken om een aantal vragen te beantwoorden die gerelateerd zijn aan het onderzoeksonderwerp en waarvoor een objectief correct antwoord al bekend is. Deze vragen dienen dezelfde complexiteit te kennen als de uiteindelijke vragen in de vragenlijst. Met behulp van deze informatie kan een optimale waarde van α worden berekend. Deze optimale α waarde kan vervolgens worden gebruikt voor het aggregeren van variabelen waarvoor de objectieve, correcte waarde niet bekend is en waarvoor onderzoekers door middel van de informantgegevens een zo nauwkeurig mogelijke groeps-waarde willen bepalen.

Op basis van de resultaten in Tabel 1 en Tabel 2 kunnen we concluderen dat het gebruiken van een wegingprocedure tot substantieel nauwkeuriger aggregatieresultaten leidt dan wanneer we simpelweg het rekenkundig gemiddelde berekenen. Verder blijkt het gebruik van gewichten die worden gebaseerd op de inschattingen van informanten van hun eigen nauwkeurigheid effectiever te zijn dan het gebruik van gewichten die worden gebaseerd op de respons van informanten. Hierbij is het niet eens noodzakelijk dat informanten voor iedere variabele waarover zij rapporteren aangeven hoeveel vertrouwen zij in de nauwkeurigheid van hun respons hebben. Een beperkter aantal metingen van het "overall" vertrouwen voor een groep vragen leidt ook tot een beperkt verlies in nauwkeurigheid terwijl het het beroep dat op informanten gedaan wordt reduceert. Dit laatste leidt mogelijk tot een verhoging van de respons en van de kwaliteit van de verzamelde gegevens.

Tenslotte liet een vervolgstudie zien dat de gebruikte methoden ook voor de aggregatie van voorspellingen van informanten tot goede resultaten leidt. Naast vertrouwenscores gebruikten we hier ook de competentie van informanten als gewichten in onze weging-procedures. Het gebruik van competentiescores als gewichten leidt tot vergelijkbare resultaten zolang het de competentie op een taak betreft die vergelijkbaar is met de taak waarvoor geaggregeerd wordt.

6. CONCLUSIE

In marktonderzoek naar organisatiegegevens zullen informanten een belangrijke bron van informatie zijn. De informatie die door informanten verstrekt wordt zal vaak niet honderd procent nauwkeurig zijn en dus een foutencomponent bevatten. Het gebruik van meerdere informanten (in plaats van één enkele informant) zal daarom de kwaliteit van responsgegevens verhogen en de validiteit van gevonden relaties tussen variabelen marktonderzoek vergroten. Omdat verschillende informanten vaak verschillende meningen zullen hebben over de waarde van organisatievariabelen is het de vraag hoe uit deze verschillende meningen één score voor een organisatie(onderdeel) moet worden afgeleid. De nadruk in dit artikel ligt op de wijze waarop responsgegevens van meerdere informanten geaggregeerd moeten worden om een zo nauwkeurige waarde voor een organisatievariabele te krijgen.

In onze onderzoekscontext leidt het berekenen van een simpel (ongewogen) rekenkundig gemiddelde al tot nauwkeuriger waarden voor organisatievariabelen (vergeleken met de inschattingen van individuele informanten) doordat de invloed van random fouten (error) zoals die aanwezig is in de informatie van individuele informanten afneemt naarmate het aantal informanten toeneemt. Het ongewogen rekenkundig gemiddelde is echter substantieel minder nauwkeurig dan gewogen gemiddelden waarin het vertrouwen van de informant in de eigen informatie of zijn/haar nauwkeurigheid met betrekking tot het verstrekken van gelijksoortige informatie, of responsgegevens als gewichten worden gebruikt. Het gebruik van deze gewichten is effectief omdat ze informatie over systematische fouten in de respons van informanten bevatten.

In tegenstelling tot onderzoek dat laat zien dat het wegen van de respons van individuen niet nodig is (zie bijvoorbeeld Armstrong 1986; Einhorn en Hogarth 1975) tonen onze resultaten dus aan dat weging van de informatie van verschillende informanten wel degelijk effectief is. Een mogelijke verklaring voor de effectiviteit van onze gewichten kan liggen in het type gewichten dat wij gebruiken. In onze aanpak weegt de informatie van meer competente informanten en informanten met meer vertrouwen zwaarder dan informatie van minder competente informanten en informanten met minder vertrouwen. Als gewichten op basis van competentie worden berekend dan is het van belang dat de competentiegewichten op basis van vergelijkbare taken wordt berekend. Competentie en expertise van informanten blijken domein-specifiek te zijn (Ashton en Ashton 1985). Hoewel het voor informanten moeilijk kan zijn om hun eigen nauwkeurigheid in te schatten en hoewel deze inschattingen zelf systematische afwijkingen kunnen kennen (Alba en Hutchinson 2000) blijken zij toch nog steeds nuttige informatie te bevatten. De aggregatiemethoden die wij in dit artikel presenteren hebben het grote voordeel dat zij zowel eenvoudig bruikbaar als effectief zijn. Onze resultaten bevestigen de bevinding van Clemen (1989) dat relatief eenvoudige aggregatiemethoden vaak opvallend effectief zijn als ze vergeleken worden met meer complexe methoden.

Voor marktonderzoekers die onderzoek onder organisaties doen hebben de resultaten van ons onderzoek een aantal *implicaties*. Allereerst is het voor de nauwkeurigheid van de verzamelde gegevens belangrijk om per organisatie niet één maar het drie informanten te ondervragen. Als vervolgens gegevens van meerdere informanten verzameld zijn dan moet niet simpelweg een gewogen gemiddelde bepaald worden om zo een organisatiescore te verkrijgen. Het is beter om een gewogen gemiddelde te berekenen waarbij

of de competentie van informanten of hun vertrouwen in hun eigen inschattingen als gewichten gebruikt wordt. Om onze methode toe te kunnen passen moeten onderzoekers informanten vragen om (1) een inschatting van de kwaliteit van hun respons te geven (bijvoorbeeld op een schaal van 1 tot 9 en (2) een aantal vragen te beantwoorden die gerelateerd zijn aan het onderzoeksonderwerp maar waarvoor een objectief correct antwoord al bekend is. Met behulp van deze laatste informatie kan de competentie van de informant worden vastgesteld. Aan de hand van de antwoorden op deze twee vragen kan een optimale waarde van α worden berekend. Deze optimale α waarde kan vervolgens worden gebruikt voor het aggregeren van variabelen waarvoor de objectieve, correcte waarde niet bekend is en waarvoor onderzoekers door middel van de informantgegevens een zo goed mogelijke waarde willen bepalen.

Al met al denken wij dat wij een aantal effectieve en gemakkelijk toepasbare aggregatiemethoden ontwikkeld hebben die het mogelijk maken om op een correctie manier informatie van meerdere informanten te aggregeren. De gepresenteerde aggregatiemethoden blijken effectief in een gesimuleerde omgeving. Het is belangrijk om ons onderzoek te repliceren in realistische praktijksituaties om zo vast te stellen hoe effectief onze methoden in dergelijke situaties zijn. Verder is het interessant en verdient het aanbeveling de waarde van onze methoden te bepalen in het geval van meer complexe variabelen (zoals de marktgerichtheid of machtspositie van organisaties) dan het soort waarover wij in ons onderzoek gegevens hebben verzameld.

LITERATUUR

- Alba, Joseph W. en J. Wesley Hutchinson, 2000. "Knowledge Calibration: What Consumers Know and What They Think They Know," *Journal of Consumer Research*, 27, September, 123-156.
- Armstrong, J. Scott, 1986. "The Ombudsman: Research on Forecasting: A Quarter-Century Review, 1960-1984," *Interfaces*, 16:1, 89-109.
- Ashton, Alison Hubbard en Robert H. Ashton, 1985. "Aggregating Subjective Forecasts: Some Empirical Results," *Management Science*, 31, 12, 1499-1508.
- Bruggen, Gerrit H. van, Gary L. Lilien en Manish Kacker, 2002. "Informants in Organizational Marketing Research: Why Use Multiple Informants and How to Aggregate Responses," *Journal of Marketing Research*, November, Forthcoming.
- Clemen, Robert T., 1989. "Combining Forecasts: A Review and Annotated Bibliography," *International Journal of Forecasting*, 559-583.
- Clemen, Robert T. en Robert L. Winkler, 1993. "Aggregating Point Estimates: A Flexible Modeling Approach," *Management Science*, 39, 4, 501-515.
- Einhorn, Hillel J. en Robin M. Hogarth, 1975. "Unit Weighting Schemes for Decision Making," *Organizational Behavior and Human Performance*, 13, 171-192.
- Ferrell, William R., 1985. "Combining Individual Judgments," in: G. Wright (ed.), *Behavioural Decision Making*, New York: Plenum Press, 111-145.
- Glazer, Rashi, Joel Steckel en Russell Winer, 1992. "Locally Rational Decision Making: The Distracting Effect of Information on Managerial Performance," *Management Science*, Vol. 38, 212-226.
- Hill, Gayle W., 1982. "Group Versus Individual Performance: Are N+1 Heads Better Than One?," *Psychological Bulletin*, 91, 3, 517-539.
- Hogarth, Robin M., 1978. "A Note on Aggregating Opinions," *Organizational Behavior and Human Performance*, 21, 40-46.

- James, Lawrence R., 1982. "Aggregation Bias in Estimates of Perceptual Agreement," *Journal of Applied Psychology*, Vol. 67, No. 2, 219-229.
- Kumar, Nirmalya, Louis W. Stern en James C. Anderson, 1993. "Conducting Interorganizational Research Using Key Informants," *Academy of Management Journal*, 36, 6, 1633-1651.
- Larréché, Jean Claude en Hubert Gatignon, 1990. *MARKSTRAT2: A Marketing Simulation Game*. Palo Alto, CA: The Scientific Press.
- Larréché, Jean Claude en Reza Moïn pour, 1983. "Managerial Judgment in Marketing: The Concept of Expertise," *Journal of Marketing Research*, Vol. XX, 110-121.
- Libby, Robert en Roger K. Blashfield, 1978. "Performance of a Composite as a Function of the Number of Judges," *Organizational Behavior and Human Performance*, 21, 121-129.
- Morris, P.A., 1977. "Combining Expert Judgments: A Bayesian Approach," *Management Science*, 23, 679-693.
- Philips, Lynn W., 1981. "Assessing Measurement Error in Key Informant Reports: A Methodological Note on Organizational Analysis in Marketing," *Journal of Marketing Research*, Vol. XVIII, November, 395-415.
- Seidler, John, 1974. "On Using Informants: A Technique for Collecting Quantitative Data and Controlling for Measurement Error in Organization Analysis," *American Sociological Review*, Vol. 39 (December), 816-831.