

De kwantitatieve betekenis van kwalitatieve uitdrukkingen van onzekerheid

Wulfert P. van den Brink* & Mario van Schaik

Samenvatting

Er is veel onderzoek gedaan naar de numerieke betekenis van verbale onzekerheidsexpressies zoals "zelden", "mogelijk" en "zeer waarschijnlijk". Dit onderzoek betreft vooral het engelse taalgebied. Uit het onderzoek komt naar voren dat niet alle expressies even geschikt zijn om onzekerheid tot uitdrukking te brengen. Voor veel van de onderzochte expressies geldt dat er tussen personen een te grote variabiliteit bestaat in de veronderstelde numerieke betekenis. Het gebruik van deze expressies kan beter vermeden worden omdat ze vager zijn dan bedoeld.

In het hier gerapporteerde onderzoek is gezocht naar een aantal nederlandse onzekerheidsexpressies die de gehele schaal van 0-100 procent bestrijken en ieder voor zich een zo eenduidig mogelijke betekenis hebben. Dit heeft geresulteerd in een voorstel voor een gestandaardiseerde lijst met uitdrukkingen waarvan de numerieke betekenis binnen vrij nauwe grenzen vastligt. Wie deze expressies gebruikt weet hoe de aangesprokene de expressie zal interpreteren.

* Universiteit van Amsterdam, Fakulteit der Psychologie, Vakgroep Methodenleer, Roetersstraat 15, 1018 WB Amsterdam.
Tel. 020 - 525.6879

Inleiding

In de communicatie tussen mensen moeten regelmatig uitspraken gedaan worden die gebaseerd zijn op onvolledige en onzekere informatie: "De bijwerking van dit medicijn is in uw geval waarschijnlijk gering"; "Er bestaat een redelijke kans dat u er met een boete vanaf komt"; "Mogelijk worden de buien vergezeld van zware windstoten". Het is in dergelijke situaties gebruikelijk om de onzekerheid verbaal in plaats van numeriek weer te geven. Zo wordt weliswaar de schijnprecisie van een puntschatting vermeden, maar het is natuurlijk de vraag wat de zender met de geverbaliseerde kansschatting bedoelt en welke boodschap de aangesprokene ontvangt. Zeker is dat met de verbale onzekerheidsuitdrukking een bewuste vaagheid wordt nagestreefd. De kennis is onvolledig en onzeker en de gebruikte verbale kansuitdrukking vormt in wezen een verbaal betrouwbaarheidsinterval. De vraag is echter wat de grenzen en het betrouwbaarheidsniveau van dit interval zijn. Men zou kunnen pleiten voor het rechtstreeks gebruik van numerieke intervallen: "De kans op bijwerkingen bij dit medicijn ligt in uw geval tussen de één en drie procent". Maar dat vereist een kennis van zaken die gewoonlijk ontbreekt. En zelfs bij de statistische verwerking van onderzoeksgegevens, waar het in veel gevallen wel mogelijk is om een betrouwbaarheidsinterval te bepalen voor het experimentele effect, kiest men bij voorkeur voor de verbale uitspraken significant versus niet significant. In inleidingen in de statistiek wordt gepleit voor het gebruik van betrouwbaarheidsintervallen omdat die inzicht verschaffen in de mogelijke grootte van het experimentele effect en daarmee in de praktische relevantie ervan. De uitspraak dat het resultaat van het onderzoek significant is, geeft het inzicht in de grootte en dus de relevantie van het effect niet. Het betekent slechts dat een onderzoeksresultaat onder de nulhypothese een overschrijdingskans bezit van vijf procent of minder en leidt op grond daarvan tot verwerping van die nulhypothese. Cowles en Davis (1982) merken bovendien op dat het vijfprocents criterium bij de significantietoets alleen zinvol is wanneer zowel wetenschappers als niet-wetenschappers het er over eens zijn dat gebeurtenissen met een kans van vijf procent of minder zodanig zeldzaam zijn dat het onderzoeksresultaat niet meer aan toeval kan worden toegeschreven.

Er bestaat dus een duidelijke voorkeur voor de nogal arbitrair gedefinieerde en weinig informatieve verbale termen significant en niet significant boven de wel informatieve numerieke betrouwbaarheidsintervallen. Een soortgelijk verschijnsel treedt op bij het gebruik van beschrijvende statistische grootheden. Correlaties worden vaak op arbitraire wijze in klassen ingedeeld die vervolgens van een verbaal etiket worden voorzien. Bijvoorbeeld: zeer hoog ($r \geq 0,90$), hoog ($0,70 \leq r < 0,90$), gemiddeld ($0,30 \leq r < 0,70$) en laag ($r < 0,30$). Heel subtiel is de indeling van Landis en Koch (1977) voor de overeenstemmingsmaat kappa: 0,00 = poor; 0,01 - 0,20 = slight; 0,21 - 0,40 = fair; 0,41 - 0,60 = moderate; 0,61 -

0,80 = substantial en 0,81 - 1,00 = almost perfect. Kennelijk bestaat er de behoefte om de puntschatting te vervangen door een verbale uitdrukking. Op zich is dat niet zo vreemd. De onzekerheid die inherent is aan de puntschatting wordt op deze wijze binnen gebracht. Maar dat gebeurt met deze etiketten op wel heel arbitraire wijze. Het berekenen van de betrouwbaarheidsintervallen voor r en κ zou een stuk informatiever zijn.

Alhoewel numerieke informatie veel nauwkeuriger is, bedienen mensen zich toch bij voorkeur van verbale termen bij het doen van onzekere uitspraken. Dit geldt zelfs bij de statistische verwerking van meetgegevens, waar andere mogelijkheden voorhanden zijn, maar natuurlijk nog sterker bij het doen van onzekere uitspraken over situaties in het dagelijks bestaan. Desgevraagd is men van mening dat verbale termen natuurlijker zijn, gemakkelijker te gebruiken, beter te begrijpen omdat men in woorden en niet in getallen denkt en de onzekerheid beter tot uitdrukking brengen (Wallsten, Budescu, Zwick en Kemp, 1993). Wel ontvangt men, vooral als het onderwerp belangrijk is, bij voorkeur numerieke informatie omdat die nauwkeuriger is (Brun en Teigen, 1988 ; Wallsten et al., 1993). Maar die numerieke informatie wordt door leken in veel gevallen niet op de juiste statistische wijze geïnterpreteerd (Brun en Teigen, 1988).

Individuele personen maken op consistente wijze gebruik van de verbale termen maar tussen personen bestaan vrij grote verschillen in het gebruik en de interpretatie van de uitdrukkingen (Beyth-Marom, 1982 ; Budescu en Wallsten, 1985 ; Mosteller en Youtz, 1990 a). Berucht is bijvoorbeeld de variatie in de veronderstelde numerieke betekenis van de term possible. Mosteller en Youtz (1990 a) vonden bij een groep wetenschapsjournalisten op een schaal van 0 tot 100 een mediane schatting van 38 met een kwartielafstand $Q = Q_3 - Q_1$ ter grootte van 43. Maar niet alle onzekerheidsuitdrukkingen zijn even vaag. Het moet mogelijk zijn een beperkte verzameling uitdrukkingen te selecteren waarvoor geldt dat iedereen het over de ordening eens is en die zonder al te veel verwarring gebruikt kunnen worden (Budescu en Wallsten, 1985). Mosteller en Youtz (1990 a) hebben een meta-analyse uitgevoerd op de literatuur over waarschijnlijkheidsuitdrukkingen om een overzicht te krijgen van goede en minder goede expressies. Een expressie is minder goed naarmate de variabiliteit in de veronderstelde normatieve betekenis groter is. Ze hebben op grond van de meta-analyse 52 uitdrukkingen geselecteerd die ieder een verschillende mate van onzekerheid representeren. Ook hebben ze de 52 uitdrukkingen voorgelegd aan een grote groep wetenschapsjournalisten om de numerieke betekenis nog eens te bepalen. Op grond van de resultaten van de meta-analyse en het onderzoek van Mosteller en Youtz heeft Kadane (1990) een voorstel gedaan voor een voorlopige gestandaardiseerde lijst van elf uitdrukkingen. Mosteller en Youtz (1990 b) hebben zich accoord verklaard met dit voorlopige voorstel. De lijst met expressies, die op enkele punten nog voor verbetering vatbaar is, staat in Tabel 1. De tabel bevat, naast de expressies, ook de mediane numerieke schattingen (Me) en de kwartiel afstanden (Q).

Tabel 1 : De gestandaardiseerde lijst van Kadane

Interval	Uitdrukking	Me	Q
0 - 5	almost never	3	3
5 - 15	seldom	10	10
15 - 25	infrequent	17	12
25 - 35	sometimes	25	17
35 - 45	less than an even chance	40	5
45 - 55	even chance	50	1
55 - 65	more often than not	60	3
65 - 75	often	72	10
78 - 85	high probability	82	10
85 - 95	very high probability	92	5
95 - 100	certain	100	1

Kadane heeft de schaal die van 0 tot 100 loopt, ingedeeld in elf deelintervallen en bij ieder deelinterval een expressie gezocht waarvan de mediaan binnen het interval valt en de kwartielfstand minimaal is.

Beroepsbeoefenaars zoals psychologen, medici en juristen moeten vaak onzekere uitspraken doen en in bijvoorbeeld persoonlijkheidsvragenlijsten komen regelmatig waarschijnlijkheidsuitdrukkingen voor. De gebruikte uitdrukkingen zullen in veel gevallen te vaag zijn en misverstanden opwekken. Het gebruik van een gestandaardiseerde lijst van uitdrukkingen kan de communicatie veel eenduidiger maken. De uitdrukkingen in een gestandaardiseerde lijst hebben het voordeel dat ze de gehele kansschaal bestrijken met minimale overlap tussen aangrenzende uitdrukkingen. Bovendien zijn voor iedere uitdrukking schattingen bekend voor de mediane numerieke betekenis en de bijbehorende kwartielfstand. Dat maakt het mogelijk om de uitdrukkingen, zo nodig, op redelijk eenduidige wijze numeriek te vertalen tot een intervalschatting.

Hamm (1991) geeft een voorbeeld van de misverstanden waartoe het ontbreken van standaardisatie kan leiden. De NASA wilde de kans op een ongeval met de space shuttle bepalen. Aan de ingenieurs werd gevraagd om de betrouwbaarheid van de verschillende space shuttle componenten met kansuitdrukkingen te omschrijven. Die uitdrukkingen werden vervolgens op min of meer willekeurige wijze numeriek vertaald. Het resultaat van de berekeningen gaf aanleiding tot een overdreven vertrouwen in de space shuttle technologie. Volgens de ingenieurskrant van 24 januari 1991 schatte men de kans op een ongeval gelijk

aan 10^{-5} . Dat betekent bij een dagelijks lancering eens in de 300 jaar een ongeval. Na de ramp met de Challenger is de kans op een ongeval bijgesteld tot $2 \cdot 10^{-2}$, een factor 2000 maal zo groot. Het gebruik van gestandaardiseerde expressies zou ongetwijfeld al in eerste instantie een meer realistische schatting voor de kans op een ongeval gegeven hebben.

In het in dit artikel te rapporteren onderzoek wordt een eerste poging gedaan om tot een Nederlandstalige gestandaardiseerde lijst van onzekerheidsexpressies te komen. Maar voordat dit onderzoek besproken kan worden, moet eerst nog aandacht besteed worden aan de invloed van de context waarin de expressies gebruikt worden op de veronderstelde numerieke betekenis van de expressies.

De variatie in de schattingen van de numerieke betekenis van een onzekerheidsexpressie kan voor een belangrijk deel worden toegeschreven aan individuele verschillen in taalgebruik (Budesu en Wallsten, 1985). Daarnaast speelt ook de context waarin de expressie gebruikt wordt een rol. In het merendeel van de onderzoeken werden de onzekerheidsuitdrukkingen zonder context aan de proefpersonen aangeboden met de opdracht numerieke schattingen aan de expressies toe te kennen. Juist deze wijze van aanbieden zou een deel van de variabiliteit in de schattingen kunnen verklaren. Mogelijk voegen de proefpersonen hun eigen context toe aan de expressie en voeren ze dus allemaal een iets andere taak uit. Wanneer deze veronderstelling juist is, zou het aanbieden van een expressie in een voor alle proefpersonen identieke context tot grotere overeenstemming over de numerieke betekenis moeten leiden. Dit blijkt echter niet het geval. Beyth-Marom (1982) en Brun en Teigen (1988) hebben laten zien dat de variatie in de schattingen meestal toeneemt wanneer alle proefpersonen de expressie in dezelfde context aangeboden krijgen. Dit wordt veroorzaakt door individueel verschillende opvattingen over de in de contexten opgenomen gebeurtenissen. Die opvattingen hebben vooral betrekking op de veronderstelde apriori kansen op de beschreven gebeurtenissen (Pepper en Prytulak, 1974; Wallsten, Fillenbaum en Cox, 1986; Newstead en Collis, 1987; Weber en Hilton, 1990). Er bestaat een positieve samenhang tussen de grootte van de veronderstelde apriori waarschijnlijkheid van de gebeurtenis en de numerieke schatting van de daaraan gekoppelde onzekerheidsexpressie. De variabiliteit in de schattingen neemt ook toe wanneer de contexten controversionele onderwerpen bevatten (Brun en Teigen, 1988) en wanneer de proefpersonen verschillend denken over de ernst van de gevolgen van de in de context verwerkte gebeurtenis (Weber en Hilton, 1990). Tenslotte kan de toegenomen variatie in de schattingen voor een deel toegeschreven worden aan de beschikbaarheidsheuristiek (Tversky en Kahneman, 1973). De kansschatting wordt beïnvloed door het gemak waarmee een proefpersoon zich voorbeelden van de in de context genoemde gebeurtenis voor de geest kan halen en daarin zullen de proefpersonen onderling verschillen.

De beschikbare kennis op het gebied van de onzekerheidsexpressies kan gebruikt worden om de communicatie van onzekerheid zo eenduidig mogelijk plaats te laten vinden. Daartoe wordt in het nu te beschrijven onderzoek een aantal nederlandstalige onzekerheids-

expressies geselecteerd waarvan de betekenis zo eng mogelijk vast ligt. Bij de selectie moet er op gelet worden dat de gekozen expressies de volledige kansschaal bestrijken, dat ze minimale variatie bezitten en dat ze zoveel mogelijk context-ongevoelig zijn. De te ontwikkelen nederlandstalige gestandaardiseerde lijst zal uit uitdrukkingen bestaan die zo goed mogelijk aan deze drie eisen voldoen.

Methode

Proefpersonen

Eerstejaars studenten psychologie namen als proefpersonen deel aan het experiment in het kader van hun verplichting gedurende 25 uren aan experimenten mee te werken.

Materiaal

Het onzekerheidsinterval, dat van nul tot honderd procent loopt, is op de door Kadane (1990) voorgestelde en in Tabel 1 weergegeven wijze in elf deelintervallen verdeeld. Op grond van de engelstalige literatuur zijn voor ieder van de elf intervallen twee expressies geselecteerd. In de eerste plaats zijn de elf expressies van Kadane uit Tabel 1 opgenomen. Daar zijn vervolgens elf expressies bijgezoekt. Voor elk interval is een expressie gekozen waarvoor op grond van meerdere onderzoeken geldt, dat de mediane puntschatting binnen het interval ligt en de kwartiel afstand zo klein mogelijk is. De 22 geselecteerde expressies zijn vervolgens in het Nederlands vertaald. Tabel 2 bevat de resultaten.

De uitdrukkingen zijn aan de proefpersonen aangeboden in vier verschillende context condities : geen context, een medische context, een juridische context en een persoonlijkheidsvragenlijst context. Voor de medische en juridische context zijn de zinnen deels ontleend aan de medische en juridische literatuur en deels zelf geconstrueerd. Alle 22 expressies komen zowel in de medische, juridische als geen-context conditie voor. Enkele voorbeelden van de gebruikte context zinnen zijn : "Het komt af en toe voor dat depressieve personen afhankelijk blijven van anti-depressiva" ; "Bepaalde symptomen van een psychose kunnen vaak worden bestreden met geneesmiddelen die op het centrale zenuwstelsel werken" ; "Bij fietsendiefstal volgt zeer zelden voorgeleiding voor de rechter-commissaris" ; "Het is waarschijnlijk dat getuigenverklaringen een vertekend beeld kunnen geven van de werkelijkheid" . Voor het volledige materiaal wordt verwezen naar van Schaik (1992).

Tabel 2 : De 22 geselecteerde onzekerheidsexpressies

Kadane expressies	interval	toegevoegde expressies
bijna nooit	0 - 5	zeer zelden
zelden	5 - 15	kleine kans
niet vaak	15 - 25	onwaarschijnlijk
soms	25 - 35	af en toe
in minder dan de helft van de gevallen	35 - 45	vaker niet dan wel
in de helft van de gevallen	45 - 55	even vaak wel als niet
vaker wel dan niet	55 - 65	in meer dan de helft van de gevallen
vaak	65 - 75	waarschijnlijk
grote waarschijnlijkheid	75 - 85	heel vaak
zeer grote waarschijnlijkheid	85 - 95	bijna altijd
zeker	95 - 100	altijd

De persoonlijkheidsvragenlijst bestaat uit 22 items die zonder wijzigingen ontleend zijn aan de nederlandse versies van de volgende vragenlijsten : ABV, DV, L&S, MMPI, NPV, PMT, PMT-K, PRF, SVL en VIS. Gegevens over deze tests kunnen gevonden worden in Evers, van Vliet-Mulder en ter Laak (1992). Voor twaalf van de 22 expressies uit Tabel 2 zijn in de tests items gevonden die deze uitdrukkingen bevatten. Aan deze twaalf items zijn tien testitems toegevoegd met andere onzekerheidsexpressies. De volledige lijst van 22 items is opgenomen in van Schaik (1992). Enkele voorbeelden van geselecteerde items zijn : "Ik heb bijna nooit last van pijn in de hartstreek of in de borst" (MMPI) ; "Als ik juf of meester iets moet vertellen, kom ik soms moeilijk uit mijn woorden" (SVL) ; "Ik maak mijn werk vaak slordig" (SVL) ; "Als ik aan iets begin, breng ik het af en toe tot een goed einde" (PMT).

Procedure

Alle proefpersonen kregen een duidelijk schriftelijke instructie met uitleg en voorbeelden. De instructie voor de geen-context conditie verschilde van die voor de drie condities met context. In de geen-context conditie werden als voorbeelden op zichzelf staande onzekerheidsexpressies gebruikt en in de context condities zinnen die onzekerheidsexpressies bevatten. Om duidelijk te maken wat van de proefpersonen verwacht werd, wordt hier een deel van de instructie voor de geen-context conditie geciteerd: "Het is de bedoeling dat u bij elk van de onderstaande kansuitdrukkingen twee beoordelingen maakt.

1. U geeft uw persoonlijk oordeel over de getalsmatige betekenis van de kansuitdrukking. U drukt dat oordeel uit in een percentage, dus een geheel getal tussen 0 en 100. Bij het voorbeeld zeer onwaarschijnlijk dient u nu met een percentage aan te geven welke frequentie naar uw mening door deze uitdrukking gesuggereerd wordt. Wanneer u bijvoorbeeld van mening bent dat zeer onwaarschijnlijk betekent in vijf procent van de gevallen, dan geeft u als antwoord 5%.
2. Andere mensen zullen het niet altijd met u eens zijn en voor dezelfde kansuitdrukking misschien een ander antwoord geven. Van u wordt gevraagd wat andere mensen onder de frequentie van de kansuitdrukking zeer onwaarschijnlijk zullen verstaan. U doet dit door een ondergrens en een bovengrens te geven voor de schattingen van andere personen. Het is niet de bedoeling dat u als grenzen 0 en 100 geeft. U moet grenzen geven waarvan u denkt dat de grote meerderheid van de antwoorden van andere personen daartussen zullen liggen. Als u vermoedt dat, uitzonderingen daargelaten, de antwoorden van andere personen voor de uitdrukking zeer onwaarschijnlijk zullen liggen tussen 2 en 10 procent, dan vult u als ondergrens 2% en als bovengrens 10% in".

Om volgorde effecten tegen te gaan werden van elk van de vier vragenlijsten tien, in volgorde van de items van elkaar verschillende versies gemaakt. Iedere bladzijde van de vragenlijst bevatte twee te beoordelen expressies. De expressies werden als volgt aangeboden:

"Overmatig drankgebruik leidt vaker wel dan niet tot leveraandoeningen

vaker wel dan niet = ... %

ondergrens = ... % bovengrens = ... % "

Binnen iedere context werden de verschillende versies van de vragenlijst op willekeurige wijze over de proefpersonen verdeeld. Het aantal proefpersonen dat de vragenlijsten invulde, bedroeg respectievelijk $n = 57$ (medische context), $n = 49$ (juridische context), $n = 49$ (persoonlijkheidsvragenlijst context) en $n = 46$ (geen-context).

Resultaten

Voor ieder van de 22 geselecteerde expressies zijn per conditie de mediaan Me , het eerste kwartiel Q_1 , het derde kwartiel Q_3 en de kwartielafstand $Q = Q_3 - Q_1$ van de puntschattingen van de proefpersonen berekend. Tevens zijn voor de gegeven intervalschattingen, per conditie en per expressie, de medianen en de kwartielafstanden van de ondergrenzen en de bovengrenzen bepaald. De resultaten zijn opgenomen in Tabel 3, waarbij M , J , P en G respectievelijk staan voor de medische-, juridische-, en persoonlijkheids-

vragenlijst- en geen-context conditie. Voor de volledige frequentieverdelingen wordt verwezen naar van Schaik (1992). Ter vergelijking zijn in de tabel ook de resultaten uit het onderzoek van Mosteller en Youtz (1990 a) opgenomen (conditie Y). Zij boden de expressies, zoals eerder vermeld, zonder context aan.

Omdat bij de meeste uitdrukkingen een aantal proefpersonen soms sterk afwijkende schattingen geven, is, in navolging van Mosteller en Youtz (1990 a) en het merendeel van de overige auteurs op het gebied van de onzekerheidsexpressies, gekozen voor de berekening van medianen en kwartielafstanden in plaats van gemiddelden en standaardafwijkingen. Vooral de gemiddelden van de expressies aan de uiteinden van de schaal ondervinden de invloed van de afwijkende schattingen omdat die slechts een kant op extreem kunnen zijn. De waarden van de medianen worden daarentegen nauwelijks beïnvloed door de extreme schattingen.

Op grond van de gegevens in Tabel 3 kan geconcludeerd worden dat er een behoorlijke mate van overeenstemming bestaat tussen de nederlandstalige en engeltalige geen-context resultaten (de condities G en Y). Alleen bij de expressie "onwaarschijnlijk" (vertaling van unlikely) ligt de mediaan systematisch buiten het beoogde interval (15,25). Bij de uitdrukking "af en toe" (vertaling van occasionally) ligt de mediaan van Mosteller en Youtz niet in het interval (25,35). De uitdrukking is desondanks voor dit interval geselecteerd op grond van het onderzoek van Hakel (1968) en Newstead en Collis (1987). In de nederlandstalige geen-context conditie valt de mediaan wel binnen het interval (25,35).

Uit Tabel 3 blijkt ook dat er een behoorlijke invloed van de context op de schattingen uitgaat. Bij de volgende expressies valt in minstens één van de drie contexten de mediaan buiten het bedoelde interval : bijna nooit, onwaarschijnlijk, soms, af en toe en vaak. Voor zes van de 22 uitdrukkingen geldt dat de kwartielafstand Q het grootst is in de geen-context conditie. Omdat er vier context condities zijn, kan op grond van toeval verwacht worden dat Q voor $\frac{1}{4} \times 22 = 5,5$ van de expressies in de geen-context conditie het grootst is. Voor de resultaten van dit onderzoek geldt dus niet dat het toevoegen van een context aanleiding geeft tot een grotere variabiliteit in de schattingen. Dit werd voor een beperkt aantal onderzochte expressies wel gevonden door Beyth-Marom (1982) en Brun en Teigen (1988).

Vergelijking van het bereik van de in van Schaik (1992) opgenomen frequentieverdelingen van de puntschattingen met het mediane bereik van de door de proefpersonen gegeven intervalschattingen laat zien dat de proefpersonen de variabiliteit in de schattingen en dus de vaagheid van de expressies sterk onderschatten. Voor bijvoorbeeld "zelden" in de medische context geldt dat de puntschattingen van 0 tot 30 lopen terwijl de mediane grenzen van de intervalschatting respectievelijk 2 en 10 zijn. Hetzelfde verschijnsel is ook gevonden door Brun en Teigen (1988).

Tabel 3 : Medianen (Me) , kwartielen (Q1 en Q3) en kwartielafstanden (Q).
M : medisch, J : juridisch, P : persoonlijkheidsvragenlijst, G : geen context,
Y : Mosteller en Youtz.

uitdrukking	conditie	puntschattingen				intervalschattingen			
		Q1	Me	Q3	Q	ondergrens		bovengrens	
bijna nooit 0 - 5	M	2	3	5	3	2	3	7	10
	J	5	10	15	10	5	6	15	15
	P	3	5	10	7	3	4	10	12
	G	4	5	11	7	2	4	12	13
	Y	1	3	5	4	-	-	-	-
zeer zelden 0 - 5	M	2	4	5	3	1	2	9	5
	J	5	5	10	5	2	4	10	8
	P	3	5	7	4	2	4	10	8
	G	2	5	10	8	2	4	10	9
	Y	1	3	5	4	-	-	-	-
zelden 5 - 15	M	3	5	10	7	2	4	10	9
	J	5	8	10	5	5	3	15	5
	P	5	8	10	5	5	3	12	5
	G	5	10	13	8	3	8	12	10
	Y	7	10	17	10	-	-	-	-
kleine kans 5 - 15	M	5	10	12	7	3	3	15	13
	J	10	15	20	10	10	10	20	15
	P	-	-	-	-	-	-	-	-
	G	8	10	16	8	5	8	15	10
	Y	8	10	20	12	-	-	-	-
niet vaak 15 - 25	M	5	10	20	15	5	9	18	20
	J	10	15	25	15	10	15	25	17
	P	8	15	25	17	10	15	25	18
	G	20	25	40	20	19	20	30	23
	Y	10	17	22	12	-	-	-	-
onwaarschijnlijk 15 - 25	M	3	10	12	9	5	8	15	18
	J	8	12	30	22	10	15	22	29
	P	-	-	-	-	-	-	-	-
	G	4	10	20	16	5	9	15	19
	Y	10	17	23	13	-	-	-	-

uitdrukking	conditie	puntschattingen				intervalschattingen			
		Q1	Me	Q3	Q	ondergrens		bovengrens	
soms 25 - 35	M	5	15	20	15	5	8	20	20
	J	15	25	35	20	15	17	32	25
	P	20	25	40	20	20	17	35	21
	G	20	30	48	28	25	30	40	31
	Y	18	25	35	17	-	-	-	-
af en toe 25 - 35	M	10	15	20	10	10	10	15	10
	J	10	15	20	10	10	10	20	15
	P	15	25	32	17	15	15	30	17
	G	20	30	46	26	22	30	40	26
	Y	13	20	28	15	-	-	-	-
in minder dan de helft van de gevallen 35 - 45	M	35	40	40	5	30	12	46	4
	J	30	40	40	10	30	11	45	9
	P	-	-	-	-	-	-	-	-
	G	30	40	40	10	30	11	45	9
	Y	40	40	45	5	-	-	-	-
vaker niet dan wel 35 - 45	M	30	40	60	30	30	35	49	30
	J	35	40	60	25	35	30	47	22
	P	-	-	-	-	-	-	-	-
	G	30	35	40	10	30	16	45	9
	Y	35	40	43	8	-	-	-	-
in de helft van de gevallen 45 - 55	M	50	50	50	0	45	6	55	9
	J	50	50	50	0	45	5	55	5
	P	-	-	-	-	-	-	-	-
	G	50	50	50	0	45	10	54	7
	Y	50	50	50	0	-	-	-	-
even vaak wel als niet 45 - 55	M	50	50	50	0	45	6	55	9
	J	50	50	50	0	45	4	55	5
	P	-	-	-	-	-	-	-	-
	G	50	50	50	0	45	6	55	10
	Y	50	50	50	0	-	-	-	-
vaker wel dan niet 55 - 65	M	60	60	70	10	55	9	70	15
	J	60	60	65	5	55	7	70	14
	P	-	-	-	-	-	-	-	-
	G	60	60	67	7	55	10	70	10
	Y	57	60	60	3	-	-	-	-
in meer dan de helft van de gevallen 55 - 65	M	55	60	60	5	52	4	70	10
	J	60	60	70	10	55	9	70	10
	P	-	-	-	-	-	-	-	-
	G	55	60	65	10	55	4	70	10
	Y	53	58	60	7	-	-	-	-

uitdrukking	conditie	puntschattingen				intervalschattingen			
						ondergrens		bovengrens	
		Q1	Me	Q3	Q	Me	Q	Me	Q
vaak 65-75	M	65	75	80	15	65	10	80	15
	J	70	75	80	10	70	10	85	15
	P	70	80	80	10	70	12	89	10
	G	70	75	80	10	70	11	85	10
	Y	65	72	75	10	-	-	-	-
waarschijnlijk 65 - 75	M	70	75	80	10	65	10	80	5
	J	50	65	75	25	58	12	75	20
	P	63	75	80	17	68	22	85	20
	G	57	62	76	19	57	22	75	21
	Y	65	70	78	13	-	-	-	-
grote waarschijnlijk- heid 75 - 85	M	65	75	80	15	70	12	85	10
	J	70	80	85	15	70	14	86	15
	P	-	-	-	-	-	-	-	-
	G	70	80	85	15	72	20	90	10
	Y	77	82	87	10	-	-	-	-
heel vaak 75 - 85	M	75	80	85	10	70	18	90	12
	J	68	80	85	17	70	15	85	16
	P	80	80	90	10	75	10	90	6
	G	80	85	85	5	75	10	90	5
	Y	75	83	90	15	-	-	-	-
zeer grote waarschijnlijk- heid 85 -95	M	85	90	95	10	85	9	95	7
	J	80	85	90	10	80	9	93	10
	P	-	-	-	-	-	-	-	-
	G	75	86	95	20	80	20	93	14
	Y	90	92	95	5	-	-	-	-
bijna altijd 85 - 95	M	80	88	90	10	80	15	95	5
	J	80	90	95	15	80	11	95	8
	P	80	85	90	10	80	13	95	5
	G	85	90	95	10	80	10	95	9
	Y	90	92	95	5	-	-	-	-
zeker 95 - 100	M	95	99	100	5	94	5	100	1
	J	88	90	100	12	85	15	100	5
	P	80	95	100	20	90	17	99	10
	G	85	95	98	13	86	13	100	5
	Y	99	100	100	1	-	-	-	-
altijd 95 - 100	M	95	100	100	5	90	5	100	1
	J	95	100	100	5	95	9	100	1
	P	95	99	100	5	90	6	100	1
	G	95	99	100	5	91	10	100	1
	Y	100	100	100	0	-	-	-	-

In de persoonlijkheidsvragenlijst context (P) komen een aantal expressies voor die niet tot de 22 geselecteerde uitdrukkingen behoren. Die expressies zijn ook in de geen-context conditie (G) aangeboden. Voor deze uitdrukkingen zijn de resultaten opgenomen in Tabel 4. Alleen de uitdrukkingen "een enkele keer", "meestal" en "zeer vaak" hebben acceptabele

Tabel 4 : Medianen (Me) , kwartielen (Q1 en Q3) en kwartielfstanden (Q) van de uitdrukkingen die alleen in de persoonlijkheidsvragenlijst context (P) en de geen-context conditie (G) voorkomen.

uitdrukking	conditie	puntschattingen				intervalschattingen			
		Q1	Me	Q3	Q	ondergrens		bovengrens	
een enkele keer	P	5	10	15	10	5	9	15	12
	G	5	10	15	10	5	8	15	12
wel eens	P	15	20	30	15	15	10	31	17
	G	24	30	44	20	25	23	40	25
nu en dan	P	15	20	32	17	15	21	33	28
	G	24	40	50	26	30	26	45	26
misschien	P	25	40	50	25	30	22	50	25
	G	27	37	50	23	27	20	49	22
regelmatig	P	50	68	75	25	60	27	75	10
	G	60	65	70	10	58	13	74	10
gewoonlijk	P	70	75	80	10	65	20	85	15
	G	60	75	85	25	60	25	82	22
over het algemeen	P	65	75	82	17	63	20	85	19
	G	60	72	80	20	64	15	80	21
doorgaans	P	70	80	85	15	68	17	87	16
	G	60	70	80	20	65	24	80	24
meestal	P	72	80	80	8	70	12	89	10
	G	75	80	85	10	71	16	86	12
zeer vaak	P	80	90	90	10	79	15	95	7
	G	80	85	95	15	77	13	90	11

eigenschappen. De overige uitdrukkingen zijn, gezien de grote waarde van Q , wel erg vaag. Expressies zoals "nu en dan" en "misschien" kunnen beter niet in items in vragenlijsten opgenomen worden omdat ze op zeer verschillende wijzen opgevat zullen worden.

Tabel 5 bevat het voorstel voor een gestandaardiseerde lijst onzekerheidsexpressies. De expressies zijn geselecteerd op grond van de in Tabel 3 vermelde resultaten.

Tabel 5 : Voorstel voor een gestandaardiseerde lijst onzekerheidsexpressies

interval	uitdrukking
0 - 5	zeer zelden
5 - 15	zelden
15 - 25	(niet vaak)
25 - 35	(soms)
35 - 45	in minder dan de helft van de gevallen
45 - 55	in de helft van de gevallen
55 - 65	in meer dan de helft van de gevallen
65 - 75	(vaak)
75 - 85	heel vaak
85 - 95	bijna altijd
95 - 100	altijd

De uitdrukkingen "niet vaak", "soms" en "vaak" zijn tussen haken gezet omdat ze slechts matig voldoen. "Niet vaak" en "soms" zijn context gevoelig en te vaag. Voor "vaak" is de mediaan aan de hoge kant. "In de helft van de gevallen" is niet vaag genoeg en kan mogelijk beter vervangen worden door "in ongeveer de helft van de gevallen". De overige uitdrukkingen voldoen goed.

De lijst expressies in Tabel 5 moet gezien worden als een eerste poging tot standaardisatie. De uitdrukkingen voor de intervallen (15,25), (25,35) en (65,75) zijn niet ideaal. Maar het is zeer de vraag of voor die intervallen betere expressies gevonden kunnen worden. De 22 uitdrukkingen in Tabel 3 zijn al geselecteerd uit een groot aantal expressies waarvoor engelstalige gegevens voor handen zijn. Ook de onlangs verschenen nederlandstalige onderzoeken van Eekhof, Mol en Pielage (1992) en Michels en Voeten (1993) geven voor de genoemde intervallen geen betere expressies. In die onderzoeken zijn respectievelijk 30 en 17 uitdrukkingen zonder context onderzocht. De auteurs geven voor iedere expressie het gemiddelde en de standaardafwijking van de puntschattingen. Dit bemoeilijkt een rechtstreekse vergelijking met de resultaten uit het hier gerapporteerde onderzoek. Voor het interval (15,25) bevat het onderzoek van Michels en Voeten (1993) geen enkele expressie. De keuze voor "niet vaak" wordt echter gesteund door de resultaten van Eekhof et al. (1992). Dit geldt ook voor de keuze van "soms" voor het interval (25,35). Voor het interval (65,75) vindt de keuze voor "vaak" zowel ondersteuning in de resultaten van Eekhof et al. (1992) als in die van Michels en Voeten (1993).

In Tabel 6 zijn voor de geen-context conditie de gegevens opgenomen voor de expressies uit Tabel 5, die ook onderzocht zijn door Eekhof et al. (1992) of door Michels en Voeten (1993). De drie verschillende onderzoeken laten voor deze expressies

Tabel 6 : De expressies uit Tabel 5 die ook voorkomen in ander nederlandstalig onderzoek. (B : van den Brink en van Schaik ; E : Eekhof et al. ; M : Michels en Voeten)

interval	uitdrukking	B		E		M	
		Me	Q	Gem.	St.dev	Gem.	St.dev
5 - 15	zelden	10	8	12	7	12	6
15 - 25	niet vaak	25	20	24	10	-	-
25 - 35	soms	30	28	26	13	26	12
65 - 75	vaak	75	10	72	11	71	13
85 - 95	bijna altijd	90	10	88	7	88	9
95 - 100	altijd	99	5	97	5	96	5

overeenkomstige resultaten zien. Voor de overige, niet in Tabel 6 opgenomen intervallen, bieden Eekhof et al. (1992) en Michels en Voeten (1993) geen betere uitdrukkingen. Er bestaat dus op grond van het andere nederlandstalige onderzoek geen aanleiding om de keuze in Tabel 5 te herzien.

Discussie

Uitspraken die gebaseerd zijn op onzekere informatie moeten een zekere vaagheid bezitten. Daarom is het gebruik van numerieke puntschattingen in dergelijke beweringen niet voor de hand liggend. Representatie van de onzekerheid met behulp van één enkel getal zou slechts ongefundeerde precisie creëren. Wel zou gebruik gemaakt kunnen worden van numerieke intervallen zoals in de zin : "de kans op succes ligt bij deze operatie tussen de 70 en 80 procent" . Maar in de inleiding is reeds besproken dat daar de nodige weerstand tegen bestaat. Bovendien is het maar de vraag of de numerieke uitspraken door de leek goed begrepen worden (Brun en Teigen, 1988). Het gebruik van verbale onzekerheidsexpressies ligt meer voor de hand. Het zijn in essentie ook schattingsintervallen en men vindt ze natuurlijker, beter te begrijpen en gemakkelijker te gebruiken (Wallsten et al., 1993). Het is wel verstandig om het gebruik te beperken tot expressies die een relatief eenduidige betekenis

bezitten. De in Tabel 5 opgenomen expressies voldoen aan deze voorwaarde. Door deze uitdrukkingen te gebruiken, wordt een te grote vaagheid in de meeste gevallen voorkomen. Alleen de expressies die behoren bij de intervallen (15,25), (25,35) en (65,75) zijn vager dan bedoeld. Maar hier is weinig aan te doen zoals ondermeer blijkt uit de meta-analyse van Mosteller en Youtz (1990 a). De voor deze drie intervallen geselecteerde expressies zijn weliswaar relatief optimaal maar men dient er zich van bewust te zijn dat ze desondanks heel verschillend geïnterpreteerd kunnen worden. De andere geselecteerde expressies voldoen redelijk tot goed. De voorgestelde verbale schaal kan gezien worden als een goed compromis tussen de weerstand tegen het gebruik van getallen en de noodzaak om tot een zo eenduidig mogelijke communicatie te komen.

Het in dit artikel besproken onderzoek is gericht op het vinden van goede onzekerheidsexpressies voor de gehele schaal van 0 tot 100 procent. De schaal is daarbij in een beperkt aantal categorieën ingedeeld. Dit betekent dat de resultaten niet van toepassing zijn op het schatten van kansen voor zeer zeldzame gebeurtenissen zoals rampen met tankers of kerncentrales. Risico-analyse is een belangrijk maar apart gebied met bijbehorende problemen, zoals het "range-frequency" model (Parducci, 1965), onderzoek (Slovic, Fischhoff en Lichtenstein, 1982) en literatuur.

De gebruikte proefpersonen, eerstejaars psychologie, zouden een bedreiging kunnen vormen voor de externe validiteit van het onderzoek. Maar uit de meta-analyse van Mosteller en Youtz (1990 a) blijkt dat voor dezelfde expressies in verschillende onderzoeken vrijwel gelijke resultaten worden gevonden ondanks het gebruik van geheel andere proefpersonen en variërende onderzoeksopzetten. De respondenten in de meta-analyse behoren tot zeer uiteenlopende groepen zoals studenten, werkenden in de gezondheidszorg, wetenschapsjournalisten, ingenieurs, leraren en militairen. De resultaten zijn desondanks robuust. Daarom kan aangenomen worden dat de resultaten die in dit onderzoek gevonden zijn bij eerstejaars psychologie eveneens generaliseerbaar zijn.

Literatuur

- Beyth-Marom, R. (1982). How probable is probable ? A numerical translation of verbal probability expressions. Journal of Forecasting , 1 , 257-269.
- Brun, W., & Teigen, K.H. (1988). Verbal probabilities : ambiguous, context-dependent, or both ? Organizational Behavior and Human Decision Processes , 41 , 390-404.
- Budescu, D.V., & Wallsten, T.S. (1985). Consistency in interpretation of probabilistic phrases. Organizational Behavior and Human Decision Processes , 36 , 391-405.
- Cowles, M., & Davis, C. (1982). On the origins of the .05 level of statistical significance. American Psychologist , 37 , 553-558.
- Eekhof, J.A.H., Mol, S.S.L., & Pielage, J.C. (1992). Is doorgaans vaker dan dikwijls ; of hoe vaak is soms ? Nederlands Tijdschrift voor Geneeskunde , 136 , 41-42.
- Evers, A., Vliet-Mulder, J.C. van, & Laak, J. ter (1992). Documentatie van tests en testresearch in Nederland . Assen : van Gorcum.
- Hakel, M.D. (1968). How often is often ? American Psychologist , 23 , 533-534.
- Hamm, R.M. (1991). Selection of verbal probabilities : a solution for some problems of verbal probability expression. Organizational Behavior and Human Decision Processes , 48 , 193-223.
- Kadane, J.B. (1990). Codifying chance. Statistical Science , 5 , 18-20.
- Landis, J.R., & Koch, G.C. (1977). The measurement of observer agreement for categorical data. Biometrics , 33 , 159-174.
- Michels, C.G.J., & Voeten, M.J.M. (1993). Hoe soms is vaak ? Tijdschrift voor Onderwijsresearch , 18 , 58-62.
- Mosteller, F., & Youtz, C. (1990 a). Quantifying probabilistic expressions. Statistical Science , 5 , 2-12.

- Mosteller, F., & Youtz, C. (1990 b). Rejoinder. Statistical Science , 5 , 32-34.
- Newstead, S.E., & Collis, J.M. (1987). Context and the interpretation of quantifiers of frequency. Ergonomics , 30 , 1447-1462.
- Parducci, A. (1965). Category judgment: A range-frequency model. Psychological Review , 72 , 407-418.
- Pepper, S., & Prytulak, L.S. (1974). Sometimes frequently means seldom : context effects in the interpretation of quantitative expressions. Journal of Research in Personality , 8 , 95-101.
- Schaik, M. van (1992). Standaardisatie van kansuitdrukkingen en de invloed van context. Intern Rapport . UvA, Fakulteit der Psychologie.
- Slovic, P., Fischhoff, B., & Lichtenstein, S. (1982). Facts versus fears: Understanding perceived risk. In: D. Kahneman, P. Slovic, en A. Tversky (Eds.). Judgement under uncertainty: Heuristics and biases. Cambridge: Cambridge University Press.
- Tversky, A. & Kahneman, D. (1973). Availability : A heuristic for judging frequency and probability. Cognitive Science, 5 , 207-232.
- Wallsten, T.S., Budescu, D.V., Zwick, R., & Kemp, S.M. (1993). Preferences and reasons for communicating probabilistic information in verbal or numerical terms. Bulletin of the Psychonomic Society , 31 , 135-138.
- Wallsten, T.S., Fillenbaum, S., & Cox, J.A. (1986). Base rate effects on the interpretations of probability and frequency expressions. Journal of Memory and Language , 25 , 571-587.
- Weber, E.U., & Hilton, D.J. (1990). Contextual effects in the interpretations of probability words: perceived base rate and severity of events. Journal of Experimental Psychology : Human Perception and Performance , 16 , 781-789.