

DE GEREEDSCHAPSKIST

SOMS IS IETS NIET WAT HET LIJKT...

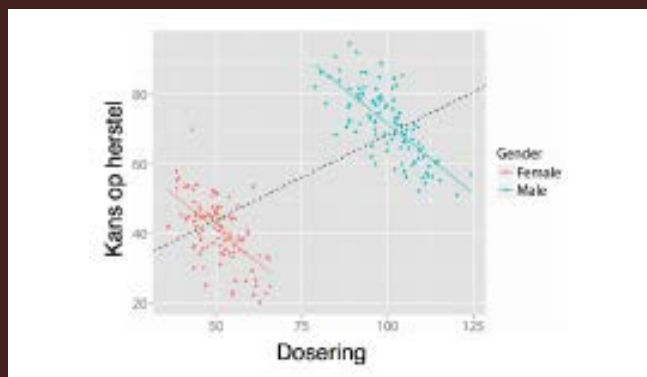
Stel: u heeft een nieuw middel ontwikkeld om psychose te behandelen. U test de effectiviteit ervan door een groep mensen met het middel te behandelen in steeds verschillende doseringen. Daarbij is de hypothese: hoe sterker de dosering, hoe beter het middel werkt. In figuur 1 vindt u de resultaten. De zwarte stippellijn met stijgende trend is het gevonden resultaat als u naar uw gehele steekproef kijkt: hoe hoger de dosering, hoe groter de kans op herstel. Op basis van deze resultaten concludeert u dat uw middel effectief is in het behandelen van psychose. Is die conclusie terecht?

Neen. Niet zonder meer. Waarom niet? Stel dat de steekproef uit twee subgroepen bestaat: mannen en vrouwen. Als we nu binnen deze subgroepen kijken naar de effectiviteit van het middel, zouden we dan dezelfde conclusie trekken als op basis van de gehele steekproef? Wederom, neen. Kijkt u als eerste naar de subgroep mannen in figuur 1 (turquoise stippen en lijn): in deze groep is de conclusie precies andersom! De lijn door de puntenwolk heen laat juist een neerwaartse trend zien. Dat betekent: hoe hoger de dosering, hoe kleiner de kans op herstel. Voor de groep vrouwen (roze stippen

'Leuker kunnen wij het als methodologen niet voor u maken, hopelijk wel makkelijker.' De rubriek De Gereedschapskist wordt verzorgd door redactieraadslid Angélique Cramer, universitair docent aan de Programmagroep Psychologische Methodenleer van de Universiteit van Amsterdam. Deze aflevering gaat over Simpsons paradox.



FOTO: JACO ROOS



FIGUUR 1. EEN VOORBEELD VAN SIMPSONS PARADOX ONTLEEND AAN FIGUUR 1 UIT KIEVIT ET AL. (2013)

en lijn) is de conclusie exact hetzelfde.

We hebben hier dus een paradox, Simpson's paradox welteverstaan – genoemd naar

de man die er in 1951 voor het eerst over publiceerde. Waar we op basis van de geaggregeerde data (mannen en vrouwen samen) concluderen dat

het middel effectief is tegen psychose, moeten we constateren dat dit binnen de subgroepen 'man' en 'vrouw' helemaal niet zo is. Hoe kan dit? In het voorbeeld is de reden wellicht dat de hogere doseringen vooral werden voorgeschreven aan de ernstigere gevallen. Die gevallen herstellen waarschijnlijk sowieso minder goed waardoor het lijkt alsof de hogere dosering geen effect sorteert.

Deze paradox is niet triviaal. In onderhavige voorbeeld kan het optreden van de paradox nogal wat consequenties hebben: behandelaars gaan het middel voorschrijven, verzekeraars vergoeden het middel ten koste van andere middelen, etc. Terwijl het middel misschien niet zo goed werkt als het lijkt!

Wat te doen? Allereerst is het belangrijk dat we ons bewust worden dat Simpson's paradox kan optreden. Daarnaast zijn er gelukkig methoden waarmee men data kan checken op het optreden van de paradox (zie Kievit, Frankenhuis, Waldorp & Borsboom, 2013). Maar dan moeten we dat voortaan dus wel gaan doen.

Kievit, R.A., Frankenhuis, W.E., Waldorp, L.J. & Borsboom, D. (2013). Simpson's paradox in psychological science: a practical guide. *Frontiers in Psychology*. doi:10.3389/fpsyg.2013.00513.