

TER DISCUSSIE

'Whiplashattributie' niet doorslaggevend voor prognose

DWALING IN STATISTISCHE INTERPRETATIE

Inge Bramsen en Pepijn Roelofs

Het beoordelen van gezondheidsstatistieken is moeilijk voor artsen, patiënten, journalisten en politici. Hierdoor trekken zij soms verkeerde conclusies, zonder dit te beseffen.¹ Dit kan onbedoelde gevolgen hebben voor de behandeling van patiënten.

Een in het oog springend actueel voorbeeld van een verkeerde interpretatie van gezondheidsstatistieken is het recente promotieonderzoek van Jan Buitenhuis dat begin juni leidde tot grote media-aandacht.² Zo kopte de *NRC*: 'Zeg niet "whiplash", zeg "spierpijn"'.³ Volgens de onderzoekers toont hun onderzoek aan dat het gebruik van de term whiplash 'doorslaggevende invloed kan hebben op de prognose'.⁴

De onderzoekers hielden echter geen rekening met de schaal waarop de onafhankelijke variabelen werden gemeten. Hierdoor klopte de door hen gemaakte vergelijking van oddsratios's (OR's) niet. Zij concluderen ten onrechte dat de 'whiplashattributie' een grotere voorspellende waarde heeft dan de ernst van de fysieke klachten. In dit artikel gaan wij in op het probleem van de verkeerde interpretatie van de statistiek in deze casus. Omdat de genoemde studie niet bij iedereen bekend zal zijn, beschrijven we eerst het onderzoek. Vervolgens geven wij de bredere context, namelijk die van de vertaling van OR's en gezondheidsrisico's naar de alledaagse medische praktijk.

DE STUDIE VAN WHIPLASH EN CAUSALE ATTRIBUTIES

Buitenhuis et al. deden een studie naar de rol van zogenaamde 'catastroferende' gedachten en causale attributies bij whiplash.⁵ 140 mensen met nekklachten na een auto-ongeluk die hiervoor een schadeclaim hadden ingediend, vulden 1 maand na het ongeluk (tijdstip 1: T1) een vragenlijst in en nog eens na 6 (T2) en na 12 maanden (T3). Voor deze studie ontwikkelden de onderzoekers de 'Causal belief questionnaire whiplash' (CBQ-W). Deze begint met de vraag: 'Mijn klachten worden veroorzaakt door ...' en somt 18 mogelijke oorzaken op. Patiënten konden antwoorden op een 4-puntsschaal ('absoluut niet', 'waarschijnlijk niet', 'waarschijnlijk wel', 'absoluut wel'). Op basis van een factoranalyse werden 4 schalen geconstrueerd: 'psychologisch' (4 items), 'ernstige verwonding' (6 items), 'wervels' (3 items), en 'spieren' (3 items). Een 5e factor met 2 items bleek niet voldoende betrouwbaar en daarom werd het item 'whiplash' als afzonderlijk item gebruikt.⁵

Causale attributie van 'whiplash' versus de inschatting van de fysieke beperkingen Twee multiple-logistische-regressieanalyses werden uitgevoerd om na te gaan welke T1-variabelen geassocieerd waren met het nog bestaan van het post-whiplashsyndroom op T2 en T3 volgens zelfrapportage van de patiënt ('ja''nee'). De onderzoekers beschrijven in hun artikel dat de causale attributie 'CBQ-W-whiplash' een onafhankelijke voorspellende waarde heeft die groter is dan die van de 'Neck disability index' (NDI). De NDI bepaalt aan de hand van zelfrapportage de ernst van de fysieke beperkingen.

Volgens ons is deze bewering onjuist, omdat de onderzoekers geen rekening hebben gehouden met de schaal waarop de onafhankelijke variabelen werden gemeten. Bij een logistische-regressieanalyse met continue onafhankelijke variabelen is de grootte van de OR afhankelijk van de schaal waarop deze variabelen zijn gemeten. Voor een zinvolle interpretatie dient de onderzoeker stil te staan bij de vraag wat een betekenisvolle verandering is op de schaal waarop een variabele is gemeten.⁶

Buitenhuis et al. vonden een OR van 3,4 voor CBQ-W-whiplash: dit betekent dat bij een toename van 1 punt op deze schaal, de waarschijnlijkheid dat iemand na 6 maanden

Hogeschool Rotterdam, Kenniskring Participatie,
Arbeid & Gezondheid, Rotterdam.

Mw. dr. I. Bramsen, methodoloog en
gezondheidszorgpsycholoog.

Drs. P.D.D.M. Roelofs, gezondheidswetenschapper
en fysiotherapeut.

Contactpersoon: dr. I. Bramsen
(i.bramsen@hro.nl).

den nog steeds het whiplashsyndroom heeft, 3,4 keer vergroot is. Deze OR is beduidend hoger dan de OR van 1,2 (exact: 1,197) voor de NDI. Echter, de NDI-scores kunnen variëren van 0-50, zodat een toename van 1 punt op deze schaal relatief klein is (1/50). De CBQ-W-whiplash loopt van 1-4 en dus is een toename van 1 punt op deze schaal verhoudingsgewijs veel groter (1/4). Dit uit zich tevens in een verhoudingsgewijs grotere OR voor deze schaal. Wanneer we voor dit verschil in het bereik van de schaal corrigeren (zie kader 'Uitleg') dan vinden we voor de NDI niet een OR van 1,2 maar van 9,5, dat wil zeggen groter dan de OR van 3,4 voor de CBQ-W-whiplash.

Deze correctie maakt duidelijk dat de voorspellende waarde van de NDI voor de prognose op de momenten T₂ en T₃ anders dan Buitenhuis et al. melden groter is dan die van de CBQ-W-whiplash. Hiermee lijkt de conclusie niet langer gerechtvaardigd dat de naamgeving van de ziekte van doorslaggevend belang is voor de prognose bij whiplash, of is deze conclusie op zijn minst discutabel. De fysieke beperkingen, zoals gemeten met de NDI, lijken voor de prognose in ieder geval belangrijker dan de genoemde causale attributie.

Er zijn, naast correctie van de OR's nog andere manieren om de relatieve bijdrage van individuele variabelen aan het logistische-regressiemodel te beoordelen. Bijvoorbeeld door inspectie van de zogenaamde 'wald-parameter'. In de voornoemde analyse was de wald-parameter voor de NDI gelijk aan 12,54 ($p < 0,001$) en voor de CBQ-W-whiplash 10,34 ($p < 0,01$). Hieraan kan men zien dat de NDI een zwaarder gewicht in de schaal legt. Ook is het mogelijk om de bijdrage van verschillende variabelen in het model te beoordelen, door het stapsgewijs op te bouwen en dan met behulp van bijvoorbeeld de 'goodness of fit'-index van Hosmer en Lemeshow te bepalen wat de bijdrage is van de in een latere stap toegevoegde variabele of variabelen.

Voorts willen we nog noemen dat het aantreffen van een statistisch significant verband tussen 2 variabelen, in dit geval de CBQ-W-whiplash en de prognose, niet automatisch een causaal verband tussen achterliggende concepten impliceert.⁷ Hiervoor zijn in de regel aanvullende aannames nodig, die voor discussie vatbaar kunnen zijn.

DE INTERPRETATIE VAN OR'S EN RELATIEVE RISICO'S.

De studie van Buitenhuis et al. illustreert dat de OR een lastig te interpreteren maat voor samenhang is.^{4,8} Vanwege de lastige interpretatie leidt het gebruik van relatieve risico's en OR's wel vaker tot een verkeerde inschatting van het belang van risicofactoren.¹ Vaak wordt verzuimd het absolute risico te vermelden, terwijl dit laatste voor een goed begrip van de uitkomsten wel van belang is. Een verdubbeling van het risico op het krijgen van een ernstige ziekte, dat wil zeggen dat het relatief

risico gelijk is aan 2, lijkt op het eerste gezicht een sterk verband, maar het maakt wel uit of het basisrisico 0,1% is of 10%. Achter relatieve risico's en OR's kunnen dus heel verschillende absolute risico's schuil gaan.

Ander voorbeeld: angst voor 3e-generatie-anticonceptiepil Een treffende illustratie hiervan is de wereldwijde 'pilangst' die in 1995 ontstond toen bekend werd dat de 3e-generatie-anticonceptiepil het risico op trombose ernstig zou verhogen. De Britse commissie voor de veiligheid van medicijnen gaf een waarschuwing dat het risico op potentieel levensbedreigende trombose door het gebruik van 3e-generatie-anticonceptiemiddelen verdubbeld werd: een toename van 100%. Deze informatie werd op grote schaal via de media verspreid en leidde tot angst en ongerustheid. Veel bezorgde vrouwen stopten onmiddellijk met de pil.

Maar welk absoluut risico ging er eigenlijk schuil achter deze berichten? Studies hadden aangetoond dat van elke 7000 vrouwen die eerder de 2e-generatiepil slikten, er ongeveer 1 trombose kreeg; dit aantal nam toe tot 2 vrouwen per 7000 die de 3e-generatiepil slikten. Het absolute risico was dus 1 op 7000 en de relatieve toename was inderdaad 100%. De pilangst leidde tot naar schatting 13.000 extra abortussen in het daaropvolgende jaar in Engeland en Wales.¹

In zijn algemeenheid moet men voorzichtig zijn met het vertalen van OR's naar aanbevelingen voor de dagelijkse praktijk. Zelfs indien een verband tussen een risicofactor en een uitkomstmaat statistisch significant is, moet dit verband zeer sterk zijn voordat het prognostische waarde heeft voor de individuele patiënt.⁹ Wil een risicofactor prognostische waarde hebben dan moet het verschil tussen de 2 groepen, bijvoorbeeld tussen wel of niet genezen zijn na 1 jaar, zo groot zijn dat de overlap in de scoreverdeling tussen de 2 groepen beperkt is.

Zeer hoge OR nodig voor belangrijke bijdrage aan diagnostische test Zelfs indien sprake is van internatio-

UITLEG

Correctie van een oddsratio (OR) voor het bereik van de schaal van de betreffende variabele De 'Neck disability index' (NDI) heeft een bereik van 0-50 en de 'Causal belief questionnaire whiplash' (CBQ-W) van 1-4. Om een OR van 1 punt in het bereik van de CBQ-W te kunnen vergelijken met een OR van 1,197 in de NDI kan men de volgende formule gebruiken: $OR^{\text{weginfactor}} = 1,197^{(50/4)} = 9,466$.⁶ Hierdoor verkrijgt men voor de NDI een OR die het effect weergeeft van een toename van 1/4 van het totale bereik van de schaal van 0-50. Deze correctie moet gezien worden als een benadering, omdat men ook rekening moet houden met de scoreverdeling die in een specifieke onderzoekspopulatie wordt aangetroffen.

naal aanvaarde risicofactoren blijkt de prognostische waarde hiervan voor de individuele patiënt soms gering.^{9,10} Dit blijkt bijvoorbeeld uit de 'Framingham heart study'.¹⁰ Deze studie identificeerde 5 biomarkers die een onafhankelijke statistisch significante bijdrage leverden aan de voorspelling van cardiovasculaire gebeurtenissen. Hiervan werd een gewogen indexscore gemaakt. Personen met een hoge indexscore hadden een 4 keer zo hoog risico op sterfte binnen 5 jaar en een verdubbeld risico op een cardiovasculaire gebeurtenis. Toch voegde deze gewogen combinatie van de topvijf van risicofactoren weinig toe aan de sensitiviteit en specificiteit van een prognostische test voor sterfte binnen 5 jaar. Dit is niet vreemd, want voor classificatie met een sensitiviteit en een specificiteit van bijvoorbeeld 0,80 is een OR van 16 nodig, en zulke sterke associaties worden zelden gevonden.

CONCLUSIE

Uit deze whiplash-casus blijkt dat het identificeren van risicofactoren, het correct interpreteren ervan en het maken van een correcte vertaling naar de dagelijkse

praktijk en de individuele patiënt een complexe zaak is. Een goed begrip van de statistiek en de beperkingen ervan is daarbij onontbeerlijk. Wij lieten zien dat men bij de interpretatie van OR's in een logistische-regressieanalyse met continue variabelen rekening dient te houden met de schaal waarop deze gemeten zijn.

Voor een goed begrip van OR's en ook van relatieve risico's is voorts van belang het absolute risico te kennen.

Tot slot impliceert een statistisch significant verband niet automatisch dat de prognostische factor ook voldoende voorspellende waarde heeft om in de klinische praktijk patiënten te classificeren en het medisch handelen hier volledig op af te stemmen.

Belangenconflict: geen gemeld. Financiële ondersteuning: geen gemeld.

Aanvaard op 13 september 2009

Citeer als: Ned Tijdschr Geneeskd. 2009;153:A865

[➤ Meer op www.ntvg.nl/opinie](http://www.ntvg.nl/opinie)

LITERATUUR

- Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S. Helping doctors and patients make sense of health statistics. *Psychological science in the public interest*. 2008;8:53-96.
- Buitenhuis J. The course of whiplash. Its psychological determinants and consequences for work disability. Proefschrift. Groningen: Rijksuniversiteit Groningen; 2009.
- Nienke Beintema. Zeg niet 'whiplash', zeg: 'spierpijn'. Alleen al diagnose 'whiplash' leidt tot veel groter risico op langdurige klachten. Interview. *NRC Handelsblad*. 2009; 4 juni.
- Buitenhuis J, de Jong PJ. De term 'whiplash' liever vermijden. Commentaar op de multidisciplinaire richtlijn voor ongecompliceerde whiplash. *Ned Tijdschr Geneeskd*. 2009;153:B2.
- Buitenhuis J, de Jong PJ, Jaspers JP, Groothoff JW. Catastrophising and causal beliefs in whiplash. *Spine*. 2008;33:2427-33.
- Hosmer DW, Lemeshow S. *Applied logistic regression*. New York: John Wiley & Sons; 1989.
- Gijn J van, Rooijmans HGM. Dwalingen in de methodologie. Causaliteit. *Ned Tijdschr Geneeskd*. 1998;142:1765-7.
- Scholten RJPM. 'Odds' en wat dies meer zij. *Ned Tijdschr Geneeskd*. 1998; 142:2452-4.
- Ware JH. The limitations of risk factors as prognostic tools. *N Engl J Med*. 2006;355:2615-2617.
- Wang TJ, Gona P, Larson MG et al. Multiple Biomarkers for the prediction of First major cardiovascular events and death. *N Engl J Med*. 2006;355: 2631-9.