

# Hoog ziekteverzuim, toeval of niet?

Isn=17479

F.H.G. Marcelissen,  
D.J. van Putten

**Verzuimcijfers vormen een belangrijke informatiebron bij het voeren van verzuimbeleid. De interpretatie van die cijfers is niet altijd gemakkelijk. Ligt het verzuim bij een afdeling een paar procent hoger dan bij de rest van het bedrijf, is dit dan reden om aandacht te schenken aan deze afdeling? Is het verschil geen toeval? Of is de samenstelling van het personeel van de afdeling verantwoordelijk voor het verschil?**

De vraag of het verschil binnen de toevalligheidsmarges valt, kan beantwoord worden door niet alleen te kijken naar het gemiddeld verzuim, maar ook naar het betrouwbaarheidsinterval. Het probleem van de ongelijke samenstelling van de onderzoeksgroep en de vergelijkingsgroep kan soms opgelost worden door (directe of indirecte) standaardisatie. Wij zullen laten zien hoe het grafisch weergeven van betrouwbaarheidsintervallen de interpretatie van verschillen (eventueel na standaardisatie) kan vereenvoudigen. Deze werkwijze is een variant op de methode van de 'grafische referentieprofielen' voor vragenlijstgegevens. Dit artikel gaat alleen in op de analyse van het verzuimpercentage. Bij het voeren van een verzuimbeleid zal ook de gemiddelde verzuimfrequentie en -duur betrokken worden. De analyse van deze laatste grootheden verloopt geheel analoog aan die van het verzuimpercentage.

## Toeval of niet?

Voorbeeld: het verzuim bij een Verkoopafdeling bedraagt 11%. De andere afdelingen van dit bedrijf blijken een verzuim te hebben van 9.5%. Is 11% nu zo hoog dat dit geen toeval kan zijn?

Deze vraag kan beantwoord worden met behulp van een statistische toets. De methode die wij hier beschrijven, namelijk een grafische weergave van de betrouwbaarheidsintervallen, maakt in principe deze statistische toetsing mogelijk, maar levert daarnaast nog andere (en wellicht veel nuttiger) informatie op. Maar allereerst een beschrijving van wat een betrouwbaarheidsinterval is, daarna zullen wij ingaan op de voordelen van deze grafische methode boven 'klassieke' toetsingsmethoden.

## Betrouwbaarheidsinterval

Het verzuimpercentage van 11%

### Berekening van het betrouwbaarheidsinterval

Een formule die rekening houdt met de "schuine" verdeling van het verzuimpercentage (het percentage ligt tussen 0% en 100%, maar meestal rond de 10%) is als volgt. Deze formule is afkomstig van Brand en Radder (1992).

$$P_l = \frac{(2lp + z^2 - 1) - z\sqrt{z^2 - (2 + 1/l) + 4p(lq + 1)}}{2(l + z^2)}$$

$$P_r = \frac{(2lp + z^2 + 1) + z\sqrt{z^2 - (2 - 1/l) + 4p(lq - 1)}}{2(l + z^2)}$$

waarbij

$P_l$  en  $P_r$  respectievelijk de linker en de rechter grens zijn van het betrouwbaarheidsinterval,

$p = VP = (\text{aantal verzuimde dagen}) / (\text{aantal kalenderdagen})$ , oftewel het verzuimpercentage

$q = 1 - VP$ , oftewel 1-het verzuimpercentage

$l = nk$

$k$  = het aantal kalenderdagen dat men gemiddeld in het huidige jaar in dienst was,

$n$  = het aantal personen in dienst

$$z = z_{1-\alpha/2} \hat{\sigma} \sqrt{k}$$

$$\hat{\sigma} = \frac{STD}{\sqrt{k}}$$

$STD$  is de standaarddeviatie van het verzuimpercentage, en

$$\sqrt{k} = \sqrt{pq}$$

Het betrouwbaarheidsinterval na standaardisatie kan op precies dezelfde wijze berekend worden als het ongestandaardiseerde betrouwbaarheidsinterval, wanneer het verzuimpercentage, de standaarddeviatie van dit percentage, en de groeps-grootte na standaardisatie bekend zijn.

$$VP_{\text{gestandaardiseerd}} = \sum n_i VP_i$$

$$STD_{\text{gestandaardiseerd}} = \sqrt{\sum \frac{n_i}{n} STD_i^2 + \sum \frac{n_i}{n} VP_i^2 - (\sum \frac{n_i}{n} VP_i)^2}$$

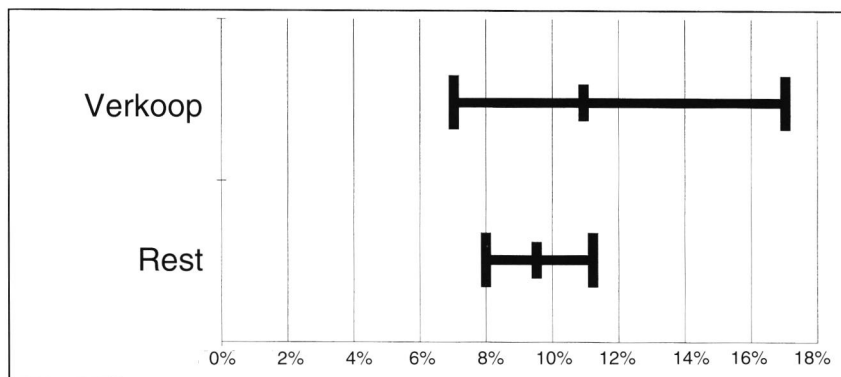
$$N_{\text{referentiegroep-na-standaardisatie}} = \frac{N_{\text{onderzoeksgroep}}}{\max(\frac{n_{\text{onderzoeksgroep}_i}}{n_{\text{referentiegroep}_i}})}$$

waarbij  $VP_{\text{gestandaardiseerd}}$ ,  $STD_{\text{gestandaardiseerd}}$  en  $N_{\text{referentiegroep}}$  na standaardisatie respectievelijk het verzuimpercentage, de standaarddeviatie behorende bij dit percentage, en de groeps-grootte zijn, allen na standaardisatie.  $VP_i$ ,  $STD_i$  en  $n_{\text{onderzoeksgroep}_i}$  zijn het verzuimpercentage, de standaarddeviatie, en de groeps-grootte van subgroep  $i$  in de onderzochte (dus niet referentie)groep.



**Tabel 1. Betrouwbaarheidsinterval van de afdeling 'verkoop' en van de rest van het bedrijf**

Afdeling	Verzuim perc	Grootte groep	Standaard deviatie	Linker betr. grens	Rechter betr. grens
Verkoop	11%	36	14.6%	7.0%	16.8%
Rest	9.5%	480	18.0%	8%	11.3%



**Figuur 1. Betrouwbaarheidsintervallen, behorende bij tabel 1**

kan worden gezien als een getal dat min op meer op basis van toeval tot stand is gekomen omdat vrijwel niemand elk jaar even vaak en even lang ziek is. 'Het werkelijk verzuim' zal ergens rond deze 11% liggen. Het betrouwbaarheidsinterval van een verzuimpercentage is nu de 'nauwkeurigheid' van dit percentage. Het is het interval waarbinnen het 'werkelijke' verzuimpercentage naar alle waarschijnlijkheid ligt. Hierbij is het gebruikelijk om een waarschijnlijkheid van 95% te hanteren (tabel 1).

Het 'werkelijk' verzuim van de afdeling Verkoop ligt met 95% zekerheid tussen de 7.0% en 16.8%. De rest van het bedrijf heeft een gemiddeld verzuim van 9.5%, met als betrouwbaarheidsgrenzen 8.0% en 11.3%. De breedte van het betrouwbaarheidsinterval van een verzuimpercentage is afhankelijk van de absolute hoogte van dit percentage, van de standaarddeviatie hiervan, en van de grootte van de onderzochte groep. In het voorbeeld is de afdeling Verkoop veel kleiner dan de rest van het bedrijf. Om deze reden is het betrouwbaarheidsinterval van Verkoop ook veel breder. Met andere woorden: hoe groter de onderzochte groep en hoe kleiner de verschillen tussen werknemers binnen de groep, des te betrouwbaarder het verzuimpercentage.

Wanneer de betrouwbaarheidsintervallen van meerdere groepen in een figuur worden weergegeven, ontstaat een profiel van betrouwbaarheidsintervallen (figuur 1).

In figuur 1 is te zien dat de betrouw-

baarheidsintervallen van de onderzoeksgroep en van de referentiegroep elkaar sterk overlappen. Dit betekent dat het verschil tussen de verzuimpercentages van beide groepen op toeval zou kunnen berusten. Er is een directe relatie tussen significantietoetsing en het profiel van de betrouwbaarheidsintervallen. Het verschil tussen twee gemiddelden is statistisch significant op 5% niveau, wanneer de 95% betrouwbaarheidsintervallen elkaar niet overlappen. Uit figuur 1 is dus af te lezen dat in het voorbeeld het verschil niet significant is. Het zou derhalve best zo kunnen zijn dat het verschil van 1.5% het volgende jaar verdwenen is.

Voor het berekenen van de betrouwbaarheidsintervallen moet men het verzuimpercentage, de standaarddeviatie hiervan, en de grootte van de onderzochte groep(en) kennen.

### Significatieniveau

Significantietoetsing is dus mogelijk met behulp van betrouwbaarheidsintervallen, maar hier ligt niet de werkelijke kracht van deze methode. Uit deze intervallen kan afgelezen worden hoe nauwkeurig het verzuimpercentage berekend is. Het vergelijken van betrouwbaarheidsintervallen laat daarom niet alleen zien hoe groot de onderlinge verschillen zijn, maar ook hoe groot de precisie is van dit verschil. Zo kan uit het grafisch weergegeven profiel meteen duidelijk zijn dat het verschil weliswaar groot is, maar toch met weinig precisie gemeten. Of andersom: het verschil is klein, maar

door de grootte van de groepen (of door kleine standaarddeviaties) toch erg nauwkeurig.

De inspectie van de betrouwbaarheidsintervallen heeft nog een ander belangrijk voordeel. Het komt vaak voor dat in één onderzoek sommige verschillen wel significant zijn, en andere niet. Dat hoeft niet te betekenen dat het verzuim van de groepen met een significant verschil het grootst is. Het is ook mogelijk dat deze afdelingen kleiner zijn, waardoor het betrouwbaarheidsinterval breder is. De grafische weergave van de betrouwbaarheidsintervallen kan daarom een veel genuanceerder beeld van de onderlinge verschillen geven dan alleen het significantpatroon.

### Standaardisatie

Het is mogelijk dat het verschil in verzuimpercentage tussen twee groepen veroorzaakt wordt door 'confounders', factoren waarin men niet geïnteresseerd is, maar die wel het verzuimpercentage beïnvloeden. Standaardisatie kan dan soms uitkomst bieden.

Standaardisatie houdt in dat bij het berekenen van de verzuimpercentages rekening wordt gehouden met de verschillende samenstelling van de groepen. Of hierdoor verschillen tussen beide groepen verdwijnen, of dat deze juist groter worden, kan vervolgens weer worden gezien in de betrouwbaarheidsintervallen.

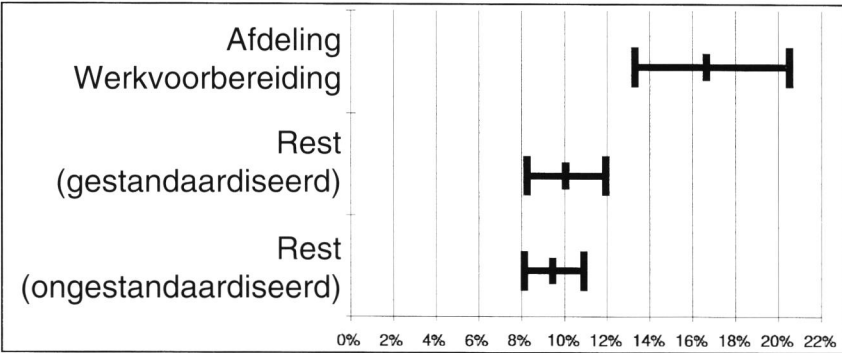
Zo bestaat in het voorbeeld van tabel 2 zowel onderzoeksgroep als referentiegroep uit mannen en vrouwen in lagere en hogere functies, maar de percentages in beide groepen verschillen. Naast standaardisatie blijkt dat het verschil tussen onderzoeksgroepen wel kleiner wordt, maar blijft bestaan. Een verschillende samenstelling verklaard dus niet de verschillende verzuimpercentages. In deze figuur is ook te zien dat het betrouwbaarheidsinterval door standaardisatie breder wordt. Met andere woorden, de mate van exactheid van het verzuimpercentage wordt kleiner. Dit effect treedt sterker op naarmate de samenstelling van beide groepen meer van elkaar verschilt. In het voorbeeld hiervoor bleek standaardisatie weinig effect te hebben op het verschil. Dit is iets dat heel vaak het geval blijkt te zijn. In een dergelijk geval kan het presenteren van gestandaardiseerde gegevens toch nut hebben: alternatieve verklaringen ('dit komt natuurlijk door ...') worden hierdoor uitgesloten.

Dit alles pleit ervoor om altijd naast de gestandaardiseerde gegevens ook de ongestandaardiseerde gegevens

Tabel 2. Betrouwbaarheidsintervallen, met en zonder indirecte standaardisatie

Referentie groep: rest van het bedrijf					
	N	verzuim %	STD verzuim %	L betr grens	R betr grens
Mannen, lagere functies	400	9.0%	18.0%	7.4%	11.0%
Mannen, hogere functies	25	7.0%	16.0%	2.8%	16.3%
Vrouwen, lagere functies	200	11.0%	19.0%	8.6%	14.0%
Vrouwen, hogere functies	45	7.0%	12.0%	4.2%	11.5%
Totaal (ongestandaardiseerd)	670	9.4%	17.9%	8.1%	10.9%
Totaal (gestandaardiseerd)	345	10.0%	18.5%	8.2%	12.2%

Onderzoeksgroep: afdeling werkvoorbereiding					
	N	verzuim %	STD verzuim %	L betr grens	R betr grens
Mannen, lagere functies	25	18.0%	15.0%	12.8%	24.7%
Mannen, hogere functies	4	4.0%	6.0%	1.0%	15.5%
Vrouwen, lagere functies	40	17.0%	15.0%	12.8%	22.3%
Vrouwen, hogere functies	0				
Totaal (ongestandaardiseerd)	69	16.6%	15.0%	13.3%	20.5%



Figuur 2. Betrouwbaarheidsintervallen behorende bij tabel 2

te laten zien. Een groot voordeel van het grafisch weergeven van betrouwbaarheidsintervallen is dat het effect van standaardisatie duidelijk te zien is, zowel op het verzuimpercentage als op de nauwkeurigheid hiervan.

Significantie?

Bij het interpreteren van het verschil tussen twee verzuimpercentages moet dus zowel worden gekeken naar de grootte van het verschil, als naar de nauwkeurigheid waarmee beide gemeten zijn. Inspectie van de betrouwbaarheidsintervallen heeft als groot voordeel boven 'klassieke' significantietoetsing dat een genuanceerde uitspraak mogelijk is: er wordt niet gesteld dat wel of geen significant verschil gevonden is, maar een verschil dat meer of minder binnen de nauwkeurigheidsmarges ligt. Vindt men bijvoorbeeld bij de evalu-

atie van een verzuimproject een afname van het verzuim dat binnen de betrouwbaarheidsmarge ligt (dus niet significant is), dan betekent dit dat het effect van het project niet 'hard' (significant) aangetoond is. Dat kan betekenen dat er geen verschil bestaat, of dat dit verschil niet aangetoond kon worden omdat de groepen te klein zijn, het verschil te klein, of de standaarddeviatie te groot. Het al dan niet significant zijn van een verschil geeft hier geen informatie over. Een grafische weergave van het betrouwbaarheidsinterval doet dit wel: het laat wel zowel de nauwkeurigheid als de 'absolute' grootte van het verschil zien.

Of relevantie?

Verschillen die statistisch niet significant zijn, of dit nu blijkt bij een significantietoets of bij vergelijking van betrouwbaarheidsintervallen, kunnen soms toch een praktische

betekenis hebben. Zo kan een afname van het verzuim met bijvoorbeeld 2% voor het bedrijf financieel veel betekenen, ook al is deze afname statistisch niet significant. Wel is er dan extra reden om rekening te houden met de mogelijkheid dat de bereikte besparing maar tijdelijk is, en dat het verzuim het volgende jaar weer op het oude niveau komt. Er kan nog een andere reden zijn om toch naar niet significante verschillen te kijken. Een afname van het verzuim met bijvoorbeeld 2% zou kunnen betekenen dat er 'in werkelijkheid' geen verschil bestaat, en dat het verzuim wellicht het volgende jaar weer op het oude niveau is. Het is ook mogelijk dat het verschil groot blijft, of zelfs toeneemt. Dergelijke trends zijn goed zichtbaar in de betrouwbaarheidsintervallen, wanneer de verzuimpercentages van meerdere jaren in één figuur worden weergegeven.

Een verschil dat te groot is voor toeval kan ook 'niet relevant' zijn. Zo is het mogelijk dat een zeer klein verschil (bijvoorbeeld 0.1%) toch statistisch significant is, wanneer de groepen maar groot genoeg zijn (enkele tienduizenden personen). Ook dit zal zichtbaar zijn in de betrouwbaarheidsintervallen. Binnen een bedrijf zal dit zich niet zo snel voordoen, maar wel wanneer bijvoorbeeld bedrijfstakken met elkaar vergeleken worden.

Literatuur

- Marcelissen, F.H.G., D.J. van Putten & J.J. Radder (1992). Graphical reference-profiles: a method for analysis and interpretation of questionnaire data. In J.L.A. van Rijkevorsel & C.C.J. Bijleveld (eds), A Reader on Applying Statistics in Public Health and Prevention. Leiden: NIPG.  
- Brand, J.P.L. & J.J. Radder (1992). Betrouwbaarheidsintervallen voor Somscores van Dichotome Items in v-Prof. Leiden: NIPG.

De auteurs

Dr. F. Marcelissen werkt als onderzoeker en programmaleider van het onderzoekswerkveld 'Ziekteverzuim, Arbeidsomgeschiktheid en Sociale Zekerheid' bij het NIA (Amsterdam).  
D.J. van Putten (arts) is als onderzoeker en coördinator van het onderzoeksveld 'Bedrijfsgezondheidszorg' verbonden aan het NIPG-TNO (Leiden).