



# Wat zeggen al die getallen eigenlijk? Over misbruik van statistiek in de processen tegen Lucia de B.

Michiel van Lambalgen



*Cognitive Science Center Amsterdam*

<http://staff.science.uva.nl/~michiell>

## Een weinig filosofisch begin

- rechtszaak betrof een verpleegster (Lucia de B.) die een ogenschijnlijk groot aantal 'incidenten' (w.o. reanimaties, sterfgevallen) meemaakte tijdens haar diensten
- bij twee overleden patienten werd een overdosis van een toxische substantie aangetroffen, maar er is geen evidentie dat Lucia deze stoffen toegediend heeft
- in de overige patienten werden geen sporen van toxische substanties gevonden, noch evidentie voor een gewelddadige dood
- behalve bij het laatste sterfgeval werd steeds een verklaring van natuurlijke dood afgegeven

## Laplace's droom: kansrekening in de rechtszaal

- Elffers, een statisticus geraadpleegd door het OM, voerde een berekening uit die zou aantonen dat onder de hypothese dat Lucia's aanwezigheid bij een incident toevallig is, de waarschijnlijkheid dat zij zoveel incidenten meemaakt kleiner is dan 1 op 342 miljoen
- in eerste aanleg (maart 2003) herformuleerde de rechtbank deze conclusie tot 'de kans dat haar aanwezigheid bij deze incidenten toeval is, is kleiner dan 1 op 342 miljoen' ; en veroordeelde haar tot levenslang
- januari 2004: de verdediging vraagt Ronald Meester en MvL naar het statistisch argument te kijken

# Dit gebruik van statistiek roept allerlei grondslagenproblemen

op ...

- gebeurtenissen met lage kans treden voortdurend op – het is niet mogelijk die te elimineren met een beroep op ‘Cournot’s principe’: ‘grote onwaarschijnlijkheid staat gelijk aan onmogelijkheid’ (zelfs nog in het standaardwerk van Kolmogorov te vinden!)
- waarom wijst een kans van 1 op 342 miljoen meer op Lucia’s schuld dan zeg een kans van 1 op 10?
- wat betekent zo’n getal eigenlijk?
- de herformulering van de rechtbank is subtiel anders dan Elffers’ oorspronkelijke formulering – maakt dat uit?
- klopt het geschetste beeld als zou het in zo’n statistisch argument alleen om berekeningen bestaan?

## Eerste waarschuwing: pas op voor getallenfetisjisme!

- de leek is geneigd om een realistische houding t.o.v. waarschijnlijkheid te hebben: een gebeurtenis *heeft* een kans zoals een baksteen gewicht heeft
- de raadsheren van het hof bleven daarom vragen: ‘als Elffers’ getal niet klopt, geef ons dan een beter getal!’
- maar dit realistische idee van kans en toeval is om meerdere redenen verkeerd:
  - er bestaat bijvoorbeeld geen welgedefinieerde waarschijnlijkheid voor de gebeurtenis dat ik voor mijn 60ste een hartaanval krijg, hoewel de gebeurtenis zelf goed gedefinieerd is
  - in *reductio ad absurdum* argumenten zoals toegepast op Lucia, is de berekende waarschijnlijkheid altijd relatief t.o.v. een kansmodel
    - daarvan kunnen er vele zijn

## Enige feiten

De verdachte werkte in twee ziekenhuizen, het Juliana kinderziekenhuis (JKZ) en het Rode Kruisziekenhuis (RKZ). De aandacht viel op de verdachte door het frequent optreden van incidenten tijdens haar diensten op het JKZ.

- JKZ: totaal 1029 diensten gedraaid door 27 verpleegkundigen; Lucia: 142
- tijdens die 1029 diensten vonden er 8 'incidenten plaats', die alle 8 tijdens een dienst van Lucia plaatsvonden
- RKZ: Lucia werkte op twee afdelingen, genummerd 41 en 42.
- RKZ-41: totaal 366 diensten: Lucia: 1
- 5 incidenten, waarvan Lucia er 1 meemaakte
- RKZ-42: totaal 339 diensten; Lucia 58
- 14 incidenten, waarvan er 5 tijdens een dienst van Lucia plaatsvonden.

## Kan dit echt toeval zijn?

- Elffers in de pers: ‘Het kan geen toeval zijn’ – nog sterker dan de formulering van de rechtbank
- maar wat is precies toeval?
- de statisticus probeert een kansmodel te formuleren om het begrip toeval exact te maken: dit is echter zelf geen wiskundige activiteit, en houdt subjectieve keuzes in
- Elffers moet bijvoorbeeld de volgende aannames maken om de intuïtieve nulhypothese in een kansmodel om te zetten waarmee gerekend kan worden
  - de verdachte heeft dezelfde kans op een incident tijdens een dienst als iedere andere verpleegster
  - er bestaat geen samenhang bestaat tussen incidenten tijdens verschillende diensten
- daarmee is toeval relatief aan een model



## Eerste poging een kansmodel op te zetten

- begin met kansen  $P(i) = p$  op een incident (bijvoorbeeld relatief aan het ziekenhuis, Den Haag, nachtdienst, ...)
- probeer te bepalen of het voorkomen van een incident *onafhankelijk* is van de aanwezigheid van Lucia:  $P(i) = P(i | L)$ , dan wel *afhankelijk*, en vermoedelijk dan  $P(i) < P(i | L)$
- om met beide hypothesen te kunnen rekenen is de *hulphypothese* nodig dat incidenten *voorwaardelijk onafhankelijk* zijn gegeven Lucia's aanwezigheid:  $P(i_1, \dots, i_n | L) = P(i_1 | L) \cdot \dots \cdot P(i_n | L)$

## Vervolg

- in het eerste geval wordt de kans om  $x$  incidenten van gegeven type in  $n$  diensten mee te maken gegeven door  $\binom{n}{x} p^x (1 - p)^{n-x}$
- in het tweede geval moet er eerst een schatting gemaakt worden van het verschil tussen  $P(i)$  en  $P(i | L)$
- een standaard statistische test vertelt je dan welke hypothese beter gesteund wordt door de data
- het probleem is echter dat er niet voldoende data zijn om de  $P(i) = p$  te bepalen
- daarom kiest Elffers een methode die  $p$  niet nodig heeft, door de sterkere nulhypothese aan te nemen dat er een *vaste* waarschijnlijkheid is dat er een incident optreedt, en dat de incidenten onafhankelijk van elkaar optreden
- deze aanname impliceert dat  $P(i | L) = P(i)$ , maar is sterker

## Elffers' 'voorwaardelijke' berekening voor het JKZ

- gegeven een urn met  $N$  ballen, waarvan er  $M$  zwart zijn. Er wordt zonder teruglegging een toevallige steekproef ter grootte van  $n$  getrokken. Wat is nu de kans dat er  $x$  ballen in de steekproef zwart zijn?

$$\frac{\binom{M}{x} \cdot \binom{N-M}{n-x}}{\binom{N}{n}}$$

- vul in
  - 'aantal ballen': aantal diensten  $N = 1029$
  - 'aantal zwarte ballen': aantal incidenten  $M = 8$
  - 'toevallige steekproef': Lucia's diensten  $n = 148$
  - 'aantal zwarte ballen in steekproef': Lucia's incidenten  $x = 8$
- de kans op de beschreven gebeurtenis is ongeveer 1 op 9 miljoen

## Elffers' 'post hoc correctie' voor het JKZ

- wat hierboven uitgerekend werd is dat *een tevoren aangewezen verpleegkundige* (nl. Lucia) alle 8 incidenten in haar diensten meemaakt – er wordt geconditioneerd op *haar* gegevens
- Elffers meent (terecht) dat dit niet eerlijk is, omdat de verdachte aangewezen is *op grond van de data*
- vergelijk: ik speel in de postcode-loterij en woon op 1052EH; de hoofdprijs valt op dat lotnummer, en iemand roept 'dat is wel erg onwaarschijnlijk!', er komt een strafrechterlijk onderzoek etc.

## Elffers' 'post hoc correctie' voor het JKZ

- een hypothese opstellen op grond van (onwaarschijnlijke) data, en die hypothese vervolgens weerleggen op grond van diezelfde data is geen kunst – de meeste gebeurtenissen 'hebben' een lage kans
- in een poging, de hypothese meer data-onafhankelijk te maken, gaat Elffers over op de hypothese dat *een* verpleegkundige op het JKZ bij toeval alle 8 incidenten meemaakt
- Elffers berekent de kans op die gebeurtenis door vermenigvuldiging met 27, wat een kans in de orde van 1 op 300.000 geeft
- maar waarom 27?

## De uitkomst: vonnis rechtbank Den Haag (maart 2003)

Onder punt 7 van de bewijsoverwegingen schrijft de rechtbank het volgende:

In zijn rapport van 29 mei 2002 komt dr. H. Elffers tot de conclusie dat de *kans* dat een verpleegkundige *bij toeval* net zoveel incidenten [...] zou meemaken als verdachte 1 op de 342 miljoen bedraagt.

Onder punt 8 vinden we:

Uit zijn rapport van 29 mei 2002 blijkt verder dat dr. H. Elffers tevens de navolgende deelkansen heeft berekend:

- a. De *kans dat bij toeval* een van de 27 verpleegkundigen bij 142 uit 1029 diensten betrokken zou zijn bij 8 incidenten... is kleiner dan 1 op 300.000
- b. De *kans dat verdachte bij toeval*...

## Vonnis rechtbank Den Haag (maart 2003)

Onder punt 11 vinden we tenslotte:

De rechtbank is van oordeel dat uit de door dr. H. Elffers uitgevoerde waarschijnlijkheidsberekeningen, zoals neergelegd in zijn rapport van 29 mei 2002, volgt dat het *uitermate onwaarschijnlijk* moet worden geacht dat de verdachte de in de tenlastelegging genoemde incidenten *bij toeval* heeft meegemaakt. Deze berekeningen geven derhalve aan dat het in hoge mate waarschijnlijk is dat er een verband bestaat tussen het werkzaam zijn van de verdachte en het zich voordoen van bedoelde incidenten.

## Grondslagenproblemen uit de kansrekening doemen op...

- de interpretatie die de rechtbank aan de berekeningen geeft is onjuist
- het 'Quine-Duhem' probleem: niet alleen de nulhypothese wordt getoetst, maar ook de hulphypotheses die gebruikt zijn om van nulhypothese naar kansmodel te komen
- het gekozen kansmodel gebruikt niet alleen (objectieve) dienstrooster data, maar is gebaseerd op allerlei subjectieve keuzes (o.a. de hulphypotheses, maar ook de post hoc correctie)
- het distilleren van een hypothese uit data om vervolgens die hypothese aan diezelfde data te toetsen is twijfelachtig
- het werken zonder exact geformuleerd alternatief voor de nulhypothese zoals bij deze 'significantie-toetsen' gebeurt, is omstreden



## Onzorgvuldige formuleringen in het vonnis

“In zijn rapport van 29 mei 2002 komt dr. H. Elffers tot de conclusie dat de *kans* dat een verpleegkundige *bij toeval* net zoveel incidenten [...] zou meemaken als verdachte 1 op de 342 miljoen bedraagt.”

- als  $H$  de nulhypothese is, en  $E$  de geobserveerde data, dan heeft

Elffers berekend: 
$$P(E | H) = \frac{1}{342 \cdot 10^{-6}}$$

- de door de rechtbank gebruikte bewering is echter  $P(H | E) =$

$$\frac{1}{342 \cdot 10^{-6}}$$

- om die conclusie te trekken is echter informatie over  $P(H)$  nodig via

$$P(H | E) = \frac{P(E | H) \cdot P(H)}{P(E)}$$

## Het 'Quine-Duhem' probleem: de rol van hulphypotheses

- Quine-Duhem: indien een meting falsificeert dat water op zeeniveau kookt bij 100C, ligt dat aan de wereld of aan de thermometer?
- Elffers was er trots op dat zijn model eenvoudig is, maar het resultaat is dat het model niet gevoelig is voor mogelijk belangrijke factoren:
  - het verschil tussen dag- en nachtdiensten (meer mensen overlijden 's nachts)
  - afhankelijkheden tussen diensten (een nog gelukte reanimatie in dienst  $n$  kan gevolgd worden door overlijden in dienst  $n + 1$ )
  - wie samen met Lucia de dienst draaide
  - of patienten Lucia uitzochten om bij te sterven ...

## Het 'Quine-Duhem' probleem: de rol van hulphypotheses

- indien de nulhypothese  $P(i) = P(i \mid L)$  verworpen wordt op basis van onwaarschijnlijke data, dan wordt de nulhypothese feitelijk verworpen samen met de hulphypotheses (zoals onafhankelijkheid van de diensten)
- bijgevolg is de onwaarschijnlijkheid van de uitkomst misschien alleen te wijten aan slecht op elkaar passen van model en werkelijkheid
- de logische structuur is:  
uit  $H \& A_1 \& \dots \& A_n$  volgt  $O$ , en  $O$  treedt niet op; wat kunnen we dan concluderen over  $H$ ?

## Kansmodellen zijn subjectieve constructen

- we hebben hierboven gezien dat er verschillende mogelijkheden zijn voor modelkeuzes
- de dronkaard onder de lantarenpaal: Elffers heeft zich bij de keuze van het model laten leiden door de *beschikbaarheid van data* – hij had ook kunnen zeggen: ‘op basis van de beschikbare data kan ik geen goed passend model maken’
- op zijn minst had het probleem in meerdere modellen bestudeerd moeten worden!
- de stem van de rede:

In applying a particular technique in a practical problem, it is vital to understand the philosophical and conceptual attitudes from which it derives if we are to be able to interpret (and appreciate the limitations of) any conclusions we draw (Vic Barnett, *Comparative Statistical Inference*, Wiley 1999).

# Retrospectief gebruik van statistiek: mag een hypothese gefalsificeerd worden door data die gebruikt zijn om de hypothese op te stellen?

- eerst een technisch punt: de post hoc correctie gebruikt niet meer alleen dienstrooster data – bedenk wat er gebeurt als de 1029 diensten over 100 verpleegkundigen verdeeld worden, met nog steeds 142 van Lucia
- dit wijst op het probleem van de *referentie klasse* – namelijk, wat is  $X$  in de volgende vraag
- ‘*wat is de kans dat een verpleegkundige uit referentie klasse  $X$  alle binnen haar afdeling optredende incidenten ook tijdens haar diensten meemaakt?*’

## Retrospectief gebruik van statistiek

- ‘wat is de kans dat een verpleegkundige uit referentie klasse  $X$  alle binnen haar afdeling optredende incidenten ook tijdens haar diensten meemaakt?’
- $X$  kan zijn: het JKZ; Den Haag; verpleegkundigen met gegeven aantal jaren IC ervaring; Europa; etc. – de getallen die daaruit komen zullen sterk verschillen
- de keuze voor een  $X$  is subjectief – het probleem is dat een singuliere gebeurtenis (hier: een concrete verpleegster die concrete incidenten meemaakt) meerdere beschrijvingen kan hebben, en dat ‘kans’ betrekking heeft op de *beschrijving*, niet op de *gebeurtenis*

## Significantie- en hypothesetoetsen

- het verwerpen van de nulhypothese zou men het liefst willen gebruiken om te concluderen dat de alternatieve hypothese  $P(i) < P(i | L)$  beter gesteund wordt door de data
- maar het model is niet in staat die hypothese uit te drukken, en de kans van de geobserveerde data onder die hypothese kan dus ook niet uitgerekend worden
- een significantietoets is een statistische test die bedoeld is om een nulhypothese te verwerpen bij een gegeven significantieniveau
- een hypothesetoets betreft ook alternatieven voor de nulhypothese in de formele beschouwing

## Significantie- en hypothesetoetsen

- binnen de psychologie heeft in 90'er jaren een methodenstrijd gewoed over de vraag of dit acceptabel is – velen meenden dat de genoemde subjectieve elementen alleen door expliciete alternatieve hypothese enigszins beheersbaar zijn
- uiteindelijk verscheen er een rapport van de APA waarin geadviseerd werd aan significantietoetsen alleen heuristische waarde toe te kennen (als opstap naar een vollediger beschrijving), maar ze niet te verbieden



## Arrest gerechtshof Den Haag

- in juni 2004 werd Lucia tot levenslang+TBS veroordeeld, ditmaal echter zonder dat de statistische argumentatie ook nog maar genoemd werd
- vanwege het criterium van ‘wettig en overtuigend bewijs’ stond het gerechtshof vrij zich te laten leiden door het idee dat deze reeks gebeurtenissen geen toeval kon zijn – dan maar zonder statistische onderbouwing, en met het voordeel dat het arrest ‘cassatie-proof’ is
- dit veronderstelt echter dat betrouwbare intuities over toeval gevormd kunnen worden zonder wiskundige precisering
- terwijl ‘toeval’ op zich nergens voor staat, en alleen betekenis heeft in de context van een model
- het onderwijs kan een belangrijke rol spelen in het zuiverder omgaan met dergelijke begrippen!
- februari 2005: advocaat mr. Stijn Franken gaat in cassatie