

Prof.dr. R.H.H. Groenwold

# Bijsluiter lezen voor gebruik



Universiteit  
Leiden

Bij ons leer je de wereld kennen

# Bijsluiter lezen voor gebruik

Oratie uitgesproken door

**Prof.dr. R.H.H. Groenwold**

bij de aanvaarding van het ambt van hoogleraar in de  
Klinische Epidemiologie, in het bijzonder methodologie van onderzoek  
naar medische behandelingen  
aan de Universiteit Leiden  
op maandag 4 maart 2019



**Universiteit  
Leiden**



*Mijnheer de Rector Magnificus, leden van de raad van bestuur van het LUMC, geachte collegae, studenten, vrienden, familie, toehoorders,*

Om te beginnen wil ik mij richten tot de studenten geneeskunde en de studenten biomedische wetenschappen in het publiek. U wordt opgeleid in een bijzondere tijd. De hoeveelheden data die momenteel worden ontsloten voor biomedisch onderzoek zijn ongekeend. Naar schatting bevat het Europese gezondheidszorgsysteem een zettabyte aan digitale informatie.<sup>1,2</sup> Een zettabyte, dat is 10 tot de macht 12 gigabyte. Als u nu in de winkel een USB-stick koopt heeft die waarschijnlijk een geheugencapaciteit van 32 gigabyte. Als we al die informatie uit de gezondheidszorg opslaan op dergelijke USB-sticks, en we knopen die aan elkaar, dan krijgen we een ketting van USB-sticks van hier naar de maan. En terug. Wat ik maar wil zeggen, dat is heel veel informatie.

Waar het verzamelen van data voor onderzoeksdoeleinden in het verleden, dan heb ik het over bijvoorbeeld 20 tot 30 jaar geleden, een majeure taak was. daar lijkt dat nu verworden tot het invullen van een aanvraagformuliertje, waarna de data schijnbaar als vanzelf in je computer verschijnen. Ik kan mij nog herinneren dat ik in het kader van mijn promotieonderzoek handmatig onderzoeksgegevens extraheerde uit een huisartsinformatiesysteem en deze overschreef op een notitieblok.<sup>3</sup> Geen voorbeeld van *good research practice*, dat geef ik toe. Tegenwoordig extraheert een datamanager met een spreekwoordelijke druk op de knop informatie over soms wel miljoenen mensen uit enorme databestanden. Daar worden weer andere databestanden aan gekoppeld waardoor een mozaïek aan informatie ontstaat, dat schijnbaar een alomvattend beeld geeft van het reilen en zeilen van de betreffende personen. Een standaard PC heeft tegenwoordig voldoende capaciteit om dergelijke omvangrijke databestanden te kunnen analyseren met allerhande software die gratis beschikbaar is via het internet. Pakketten als R of Python faciliteren het gebruik van algoritmes met beloftevolle namen als machine learning, deep

learning en artificial intelligence. En als het niet zelf lukt om deze algoritmes aan de praat te krijgen, dan is er wel een YouTube video waarin dat in gemakkelijke stappen wordt uitgelegd. Kortom, het huidige cohort studenten wordt opgeleid in een wereld met ongekeende hoeveelheden data en een toenemend aantal geavanceerde analysetechnieken. Dat biedt kansen én uitdagingen.

In deze rede zal ik proberen het onderzoek dat wordt uitgevoerd met dergelijke databestanden van een methodologische bijsluiter te voorzien. Net zoals dat geldt voor een bijsluiter bij geneesmiddelen bevat mijn methodologische bijsluiter gebruiksadviezen en beschrijft het de werkzame stof en mogelijke bijwerkingen. In het eerste deel van deze rede zal ik in gaan op de potentie die deze enorme databestanden hebben. Wat kunnen we er mogelijk mee en waarom staat er vrijwel dagelijks iets in de media over deze zogenaamde big data.<sup>4,7</sup> Of, in termen van mijn bijsluiter, ik zal de werkzame stof en de wijze van gebruik met u bespreken. Vervolgens zal ik in het tweede deel van deze rede ingaan op onjuist gebruik en overdosering. Daarbij zal ik ook stil staan bij manieren om onjuist gebruik te voorkomen of te verhelpen. In het derde deel waarschuw ik voor de bijwerkingen van onjuist gebruik. En ik zal nu maar vast verklappen dat hoe groot, of hoe representatief een databestand ook is: bij onjuiste dosering of onjuist gebruik zullen resultaten slechts leiden tot verwarring en tot dwaalsporen die uiteindelijk schadelijk zijn voor de geloofwaardigheid van de academische wetenschap. Beste toehoorders, ik hoop dan ook dat u mijn bijsluiter ter harte neemt. Laten we beginnen met het bespreken van de werkzame stof.

### **De werkzame stof**

Stel, ik heb een vaas, een heel grote vaas, die gevuld is met heel veel ballen, rode en groene ballen, misschien wel 7 miljard ballen. Nu wil ik een inschatting maken van de proportie rode en groene ballen. Wat kan ik doen? Ik kan de vaas leeg storten, en alle ballen tellen. Een ondoenlijke klus. Ik kan ook mijn arm in de vaas steken, eens flink grabbelen en van een handvol

ballen het aandeel rode en groene ballen bepalen. Stel dat ik 10 ballen heb gegrabbeld, waarvan 5 rode. De beste schatting die ik dan kan maken van de proportie rode ballen in de vaas is 0,5. Maar is dat ook de proportie in de vaas? Zou het misschien ook wel iets meer of iets minder kunnen zijn, bijvoorbeeld 0,48, of 0,53? Hoe zeker of hoe onzeker zijn we over die 0,5? Stel dat ik nog een paar grepen neem uit de vaas. In totaal haal ik er wel 100 ballen uit, waarvan er 50 rood blijken te zijn. Opnieuw schat ik de proportie rode ballen in de vaas: 0,5. Ik zal nu meer overtuigd zijn dat de werkelijke proportie inderdaad in de buurt van 0,5 ligt. De kracht van grote databestanden schuilt in dit fenomeen. Naarmate de steekproef, of onderzoeksgroep, groter wordt, neemt de onzekerheid over hetgeen ik wil schatten af.<sup>8</sup>

Waarom heb ik het vandaag met u over grote databestanden en over het grabbelen van ballen? Dat heeft alles te maken met mijn vakgebied. Dat is de klinische epidemiologie, een kwantitatieve onderzoeksdiscipline die probeert te schatten hoe vaak ziektes vóórkomen en wat de effecten zijn van bijvoorbeeld risicofactoren en van medische behandelingen.<sup>9,10</sup> Schattingen kennen onzekerheid en één van de manieren om die onzekerheid te verkleinen is grotere databestanden te gebruiken voor epidemiologisch onderzoek.

Voorbeelden van het gebruik van grote databestanden in epidemiologisch onderzoek zijn legio. Een mooi voorbeeld is de Britse Clinical Practice Research Datalink (afgekort CPRD), een databestand op basis van elektronische patiëntendossier, met informatie over meer dan tien miljoen Britten dat wordt bijgehouden sinds midden jaren 80.<sup>11</sup> Alle contacten die deze mensen hebben met hun huisarts worden geregistreerd in dit systeem, zonder dat het voor onderzoekers herleidbaar is naar individuen: de privacy is gewaarborgd. Samen met onder meer de hooggeleerde Klungel heb ik in de afgelopen jaren verschillende onderzoeken uitgevoerd in dit databestand.<sup>12-14</sup> Ook in de regio Leiden zijn verschillende van dergelijke initiatieven. Een mooi voorbeeld is het project Haagse Vaten.<sup>15</sup> Daarin wordt informatie samengebracht uit de eerste en tweede lijn uit

de regio Den Haag. Het uiteindelijke doel is het kunnen doen van onderzoek naar oorzaken van gezondheidsverschillen tussen bevolkingsgroepen en naar verschillen in prognose na bijvoorbeeld het doormaken van een hartinfarct.

Wat deze voorbeelden met elkaar gemeen hebben is de omvang van de databestanden en dat we verwachten dat ze een goede afspiegeling zijn van de dagelijkse praktijk, dat wil zeggen dat de data representatief zijn voor die dagelijkse praktijk. Juist als we onderzoek willen doen waarvan de resultaten uiteindelijk ten goede komen aan beslissingen en processen in de dagelijkse praktijk, is dat laatste uiteraard van belang. Een voorbeeld. Een Nederlandse huisarts ziet elke week gemiddeld 2 patiënten met een schimmelinfectie van de huid, een zogenaamde ‘kleine kwaal’, maar een veelvoorkomend probleem.<sup>16</sup> Als we uitspraken willen doen over deze groep patiënten die doorgaans in de eerste lijn wordt behandeld, zal een evaluatie van bijvoorbeeld alleen diegenen die zijn doorverwezen naar de tweede lijn ons niet het antwoord geven dat we zoeken.<sup>17</sup> Zinvolle resultaten komen van onderzoek met gegevens uit de eerste lijn.

Al jaren zie je dat steeds meer routinematig verzamelde zorggegevens bij elkaar worden gebracht uit bijvoorbeeld eerste en tweede lijn, gegevens over geneesmiddelen die zijn afgegeven door apothekers, of declaratiegegevens van zorgverzekeraars. Op deze manier ontstaan omvangrijke databestanden.<sup>18</sup> De werkzame stof van deze databestanden is zijn omvang en zijn representativiteit.

Nu weer terug naar die vaas met ballen. Stelt u zich nu die vaas met ballen voor als een grote vaas met allemaal Britten er in. In plaats van een paar ballen (of Britten) uit de vaas te trekken, stelt de werkzame stof uit de CPRD dataset mij in staat om relatief eenvoudig een enorme greep te nemen en – met behulp van slimme algoritmes – te tellen; de computer doet zijn werk wel. Als ik bijvoorbeeld geïnteresseerd ben in de proportie mannen in het Verenigd Koninkrijk, dan kan ik dat heel precies schatten.

### Wijze van gebruik

Waarvoor kunnen we deze enorme databestanden inzetten en op welke manier kunnen we ze gebruiken voor epidemiologisch onderzoek? Doorgaans wordt onderscheid gemaakt tussen enerzijds onderzoek dat tot doel heeft ziekte en gezondheid te voorspellen en anderzijds onderzoek dat tot doel heeft ziekte en gezondheid te begrijpen. Beide doelen zijn interessant en relevant. Enerzijds worden we steeds beter in het voorspellen van bijvoorbeeld een individuele prognose na het doormaken van een hartinfarct, anderzijds blijven we ook geïnteresseerd in het begrijpen van oorzaken die leiden tot bijvoorbeeld een recidief hartinfarct.<sup>19</sup> Onderdeel van dat begrip is inzicht in pathofysiologische processen waarop we mogelijk kunnen ingrijpen met geneesmiddelen of andere medische behandelingen.

Het evalueren van effecten van medische behandelingen heeft, mijns inziens, niet alleen tot doel om te weten hoe goed een geneesmiddel werkt, maar ook om te begrijpen hoe en waarom die behandelingen ingrijpen op het ziekteproces. Een belangrijk onderdeel daarvan is het onderscheid tussen twee (of meer) hypothetische situaties. Wat zou er gebeuren als we *deze* behandeling geven en wat zou er gebeuren als we een *andere* behandeling geven, of geen behandeling. Is die gebeurtenis voorspelbaar?

Ik heb hier een pingpongbal. Als ik deze pingpongbal nu weg werp, zal deze uiteindelijk de grond (of een van de toehoorders) raken. De pingpongbal blijft niet zweven, gaat niet de deur uit om een rondje over het Rapenburg te vliegen. Welk traject ze precies zal volgen is een functie van werpsnelheid, draaimoment, afwerphoek, luchtweerstand, de zwaartekrachtconstante, en zo nog een paar parameters. Een complex model misschien, maar als ik al die variabelen ken, kan ik heel accurate voorspellingen doen. Sterker nog, ik kan de bal zodanig beïnvloeden dat ik kan bepalen waar ze terecht komt, telkens weer. Ik kan het proces begrijpen, beïnvloeden en – gegeven de verschillende randvoorwaarden – ligt de uitkomst vast. Hoe meer randvoorwaarden ik ken, hoe meer het proces dat ik beschrijf deterministisch is vastgelegd.

Nu naar de effecten van medische behandelingen, liggen die deterministisch vast? Als er universele natuurwetten zouden bestaan voor de effecten van medische behandelingen zouden we die effecten accuraat kunnen voorspellen. Of er dergelijk natuurwetten zouden kunnen bestaan voor ziekte en gezondheid wil ik in het midden laten. Wat mij betreft zijn we momenteel nog niet in staat processen van ziekte en gezondheid volledig te vatten in cijfermatige en accurate wetmatigheden en ik zie daar in de nabije toekomst geen wezenlijke veranderingen in komen. En daarom moeten we die effecten schatten in grote, representatieve groepen van potentiële gebruikers van die behandelingen. De omvangrijke databestanden van routinematig verzamelde zorggegevens die ik u eerder heb geschetst bieden ons wellicht de mogelijkheid om dat te doen.

Ik hoop dat ik u – op dit moment in mijn verhaal – enthousiast heb kunnen maken voor het gebruiken van grote databestanden van routinematig verzamelde zorggegevens voor onderzoek naar de effecten van medische behandelingen. De werkzame stof en de wijze van gebruik zijn geschetst. Nu komen we tot het tweede deel van mijn betoog waarin ik het zal hebben over onjuist of oneigenlijk gebruik en overdosering. Bij oneigenlijk gebruik gaat het om onderzoek waarvan op voorhand kan worden verwacht dat het goede antwoord op de onderzoeksvraag niet aanwezig is in de databestanden. Onder overdosering versta ik dan niet het gebruik van te veel data, maar het gebruiken van data voor te veel doeleinden, het te vaak gebruiken van data.

### Oneigenlijk gebruik

Laten we beginnen met oneigenlijk gebruik. Gaat u in gedachten nog even mee naar die vaas met rode en groene ballen. We wilden de proportie rode ballen weten en op basis van een handvol ballen probeerden we deze proportie te schatten. Maar stel nu eens dat in onze denkbeeldige vaas de groene ballen veel zwaarder zijn dan de rode ballen, waardoor alle rode ballen bovenin de vaas komen te zitten. Of stel dat ik toevallig rode verf aan mijn grabbelhand heb. Hoeveel ballen

ik ook grabbel, mijn inschatting van de proportie rode ballen is verkeerd. En wat is de waarde van een precieze, maar onjuiste, schatting? Een systematische vertekening van de kwantiteit die ik wil schatten wordt bias genoemd. In het voorbeeld van de vaas zijn er twee mogelijke oorzaken van bias: het verschil in gewicht tussen de rode en groene ballen en de verf aan mijn hand. In epidemiologisch onderzoek hebben we dezelfde twee problemen, plus nog één.<sup>20</sup> Ik zal betogen dat deze problemen ook gezien kunnen worden als uitdagingen, waarvoor soms oplossingen mogelijk zijn. Zijn deze oplossingen er niet, dan is er sprake van oneigenlijk gebruik en is het beter om het middel (of de data) niet te gebruiken.

#### *Ontbrekende waarnemingen*

Het eerste probleem is dat databestanden, zoals het eerder genoemde elektronisch patiëntendossier, vaak onvolledig zijn. Informatie wordt verzameld om zorgprocessen te ondersteunen: van een sportieve jongedame die haar huisarts bezoekt in verband met een ingegroeide teennagel zal doorgaans niet de bloeddruk worden geregistreerd, terwijl dat wél zo zal zijn voor een zeventigjarige man met overgewicht en kortademigheidsklachten. Dit leidt tot selectief registreren, of tot het selectief ontbreken van informatie; we spreken dan vaak van ontbrekende waarnemingen. Vanuit praktisch medisch oogpunt is het triviaal dat bloeddruk alleen wordt gemeten (en dus geregistreerd) als daar een aanleiding voor is. Voor medisch onderzoek leidt het daarentegen tot grote problemen.

Stel dat we geïnteresseerd zijn in de gemiddelde bloeddruk van Nederlandse volwassenen, maar we hebben alleen bloeddrukmetingen van personen bij wie er een reden was om die bloeddruk te meten (waarschijnlijk omdat werd gedacht dat die bloeddruk te hoog was). De normale bloeddrukken ontbreken en wanneer de gemiddelde bloeddruk wordt gebaseerd op de beschikbare metingen zal dat leiden tot een overschatting van de gemiddelde bloeddruk van Nederlandse volwassenen.<sup>21</sup> Het eenvoudigweg negeren van ontbrekende waarnemingen is daarom vaak geen goede oplossing,<sup>22</sup> maar rekening houden

met de ontbrekende informatie is – door het selectieve karakter ervan – ook geen sinecure.<sup>23</sup> Immers, hoe weet je nou, wat je niet weet?

#### *Meetfouten*

Het tweede probleem is de accuraatheid van routinematig verzamelde gegevens.<sup>24</sup> Meetfouten zijn soms eenvoudig te herkennen. Een systolische bloeddruk van 1500 mmHg, of een lengte van een volwassen man van 173 m zijn evident onjuist. Deze onwaarschijnlijke waardes worden hopelijk opgepikt in een standaardevaluatie van de kwaliteit van een dataset.

Andersom geldt dat wat een meetfout lijkt, geen meetfout hoeft te zijn. Toen ik naar een kleermaker ging om mijn maten te laten opmeten voor deze toga, bleek dat mijn rechterarm 1 cm langer is dan mijn linker. “Ah, een meetfoutje,” zei ik joviaal. Dat kon de vakman niet waarderen. Het schijnt dat bij veel mensen de dominante arm iets langer is dan de niet dominante arm. Bij confectiepakken wordt daar niets mee gedaan, maar kijkt u eens om u heen: allemaal gewaden waar de ene mouw waarschijnlijk langer is dan de andere. Misschien kunt u straks bij de borrel de proef op de som nemen. Wat ik maar wil zeggen: een onverwachte bevinding is niet noodzakelijkerwijs een meetfout.

Veel subtieler dan evidente fouten zijn de meetfouten die leiden tot realistische getallen. Of meetfouten die niet van invloed zullen zijn in de dagelijkse praktijk, maar wel een rol spelen in onderzoek. De vraag “Wat weegt u ongeveer?” kan in de dagelijkse praktijk voldoende zijn om beslissingen te nemen. Dat betekent nog niet dat een dergelijke grove meting geschikt is voor medisch onderzoek. Het eigen gewicht wordt bovendien stelselmatig te laag gerapporteerd.<sup>25</sup> Bedenk dan dat er bij zwaardere mensen meer ruimte is voor onderschatting en een methodologisch probleem is geboren.<sup>26</sup> Daarbovenop speelt dan nog dat de mogelijk invloed van meetfouten verschilt wanneer we gewicht zien als aan risicofactor voor ziekte of als een uiting van een gezondheidstoestand.<sup>27</sup>

Sommige routinematig verzamelde gegevens lijken erg betrouwbaar, maar zijn slechts een grove afspiegeling van datgene waar we werkelijk in geïnteresseerd zijn. Een voorbeeld is informatie over het gebruik van een bepaald geneesmiddel. In databestanden zoals de eerdergenoemde CPRD is informatie beschikbaar over recepten die zijn uitgeschreven door huisartsen. Of de voorgeschreven geneesmiddelen ook zijn opgehaald bij de apotheek, dat weten we niet. Net zo min weten we of de patiënten – als ze het middel hebben opgehaald – het zijn gaan gebruiken en of ze het bleven gebruiken. Het is afhankelijk van het geneesmiddel en de indicatie, maar er kan wel 50% verschil zitten tussen voorschrijven en gebruik.<sup>28,29</sup>

#### *Frequentie van meten*

Deze eerste twee uitdagingen, te weten ontbrekende waarnemingen en meetfouten, komen vaak tegelijk voor, bijvoorbeeld als we kijken naar de frequentie van metingen. Een cruciaal onderdeel van epidemiologisch onderzoek is namelijk niet alleen of metingen correct zijn uitgevoerd, maar ook wanneer de metingen hebben plaatsgevonden.<sup>30</sup> Zaken die niet of nauwelijks veranderen (denk aan bloedgroep), die hoeven uiteraard maar eenmalig te worden gemeten. Veranderlijke zaken vaker. Hoe vaak, dat hangt weer af van hoe veranderlijk het proces is dat wordt bestudeerd. Rookgedrag verandert weliswaar tijdens een leven, maar in de meeste gevallen niet vaak genoeg om een meetfrequentie van één keer per week te rechtvaardigen. Hetzelfde geldt voor het optreden van chronische ziekten zoals diabetes of hartfalen. Het is doorgaans niet zo dat die ziekten binnen een tijdsbestek van een paar dagen ontstaan, of binnen een paar dagen verdwenen zijn. Kijkt een onderzoeker daarentegen naar bovenste luchtweginfecties bij kinderen, dan is een zekere fijnmazigheid natuurlijk wel vereist.

In elektronische patiëntendossiers hangt de frequentie van meten vaak af van de ernst van de ziekte die een patiënt onder de leden heeft.<sup>31</sup> Immers, tegen de ouders van een kind met een relatief onschuldige aandoening als xerosis cutis – een droge huid – zeg je dat ze terug mogen komen als het niet beter

wordt nadat ze het kind eens flink in de zalf hebben gezet. Daarentegen worden vaste controlemomenten afgesproken voor een 70-jarige patiënt die rookt, een fors verhoogde bloeddruk én een verhoogd cholesterolgehalte heeft (allemaal risicofactoren voor cardiovasculair problemen binnen afzienbare tijd).<sup>32</sup>

Als we van een afstandje kijken naar het elektronisch patiëntendossier is de frequentie van metingen informatief.<sup>31</sup> Hoe vaker iemand terug komt, hoe ernstiger zijn of haar conditie. In mijn eigen onderzoek naar de effecten van de griepvrikk, samen met de hooggeleerden Hoes en Hak, vond ik dat het aantal contacten met een huisarts in de 12 maanden voorafgaand aan het moment van vaccineren een enorm sterke voorspeller was voor overlijden tijdens het griepseizoen.<sup>33,34</sup> Maar achteraf is het makkelijk praten. Kunnen we deze kennis ook inzetten om patiënten met een verhoogt risico te identificeren? U kunt zich vast wel voorstellen dat er op de achtergrond in een huisartsinformatiesysteem een programmaatje draait dat een signaal af geeft als een patiënt bijvoorbeeld voor de 50<sup>e</sup> keer in een jaar tijd een afspraak maakt. Let op, hoog-risico-patiënt! Is dit zinvol? Welnee, de huisarts ziet die patiënt elke week, alsof ze dat zelf niet weet. Evenzo zal de inschatting van het risico van een patiënt die nooit op het spreekuur komt aanvankelijk laag zijn. En als we een iets geavanceerder risicomodel willen ontwikkelen? Bijvoorbeeld een model om het cardiovasculaire risico in te schatten. Als er jarenlang geen cholesterol is gemeten, is het cardiovasculaire risico waarschijnlijk heel laag ingeschat en daarom werd het cholesterol niet gemeten:<sup>32</sup> een ontbrekende cholesterolmeting is dus informatief! In dit geval heb je een gunstiger prognose als de meting ontbreekt. Eureka! We maken een cardiovasculair risicomodel dat er als volgt uit ziet: hoe lager het cholesterol hoe gunstiger de prognose en het ontbreken van een cholesterolmeting is helemaal gunstig. En nu stuiten we (in ieder geval in gedachten) op onjuist gebruik van routinematig verzamelde zorggegevens voor onderzoeksdoeleinden. Laten we de tijd even snel vooruit spoelen: het voorspelmodel is inmiddels geïmplementeerd in de dagelijkse



praktijk. Er zit een patiënt in de spreekkamer en we willen een inschatting maken van zijn cardiovasculaire risico. Waarschijnlijk is het beste voor de patiënt om zijn cholesterol niet te meten. In dat geval loopt hij, volgens het model, immers het laagste risico. Op het moment dat we meten neemt dat risico al toe, waarschijnlijk zelfs ongeacht wat de uitslag van die meting is. Dat is toch te gek. Kunnen we zo gemakkelijk de prognose van deze patiënt beïnvloeden? Dat lijkt een stuk effectiever dan het voorschrijven van een cholesterolverlager of een bewegingsprogramma, om nog maar te zwijgen van een leefstijladvies of een zelfmanagementapp. Ik hoop dat u een zeker ongemak voelt bij dit voorbeeld. De reden voor dit ongemak is het derde probleem.

### *Confounding*

Een derde mogelijke bijwerking is namelijk het verwarren van een waargenomen relatie met een oorzakelijk verband. Er is bijvoorbeeld een relatie tussen de hoeveelheid erwtensoep die wordt geconsumeerd en het risico op een ruptuur van een aneurysma van de buikslagader.<sup>35</sup> Is dit een oorzakelijke relatie? Dat wil zeggen, als we erwtensoep bij de wet zouden verbieden of iedereen op een strikt dieet zouden zetten van drie koppen erwtensoep per dag, zouden we daarmee dan het aantal aneurysmarupturen kunnen beïnvloeden? Nee, zowel de consumptie van erwtensoep als het risico op een aneurysmaruptuur zijn afhankelijk van iets anders, namelijk omgevingsinvloeden zoals temperatuur, guur winterweer, noem maar op. Dit fenomeen noemen we *confounding*: een waargenomen relatie tussen twee factoren is te verklaren door een gezamenlijk oorzaak van beide factoren.<sup>36</sup> Er is wel een relatie, maar dat is geen oorzakelijke. Het eten van erwtensoep hangt samen met het risico op een aneurysmaruptuur, maar is er niet de oorzaak van. Het probleem van *confounding* is dat we geneigd zijn een waargenomen relatie tussen twee fenomenen te duiden als een oorzakelijk verband.

Om te zien of er sprake zou kunnen zijn van een causaal verband tussen blootstelling aan een geneesmiddel en een bepaal-

de gezondheidsuitkomst, zouden we eenvoudigweg groepen mensen kunnen vergelijken die het ene of het andere geneesmiddel krijgen. Echter, de groepen die we dan vergelijken, bijvoorbeeld patiënten met een verhoogde bloeddruk die óf een diureticum óf een ACE-remmer krijgen voorgeschreven, zullen niet alleen verschillen ten aanzien van het gebruikte geneesmiddel, maar ook in de redenen waarom ze juist dat middel krijgen voorgeschreven, redenen zoals co-morbiditeit of onvoldoende respons op een ander bloeddrukverlagend middel.<sup>37,38</sup> Het resultaat is dat de groepen niet direct te vergelijken zijn en een waargenomen verschil in bijvoorbeeld bloeddrukdaling tussen de groepen onterecht wordt toegeschreven aan een van de twee geneesmiddelen – een voorbeeld van *confounding*. De onvergelijkbaarheid van onderzoeksgroepen is een belangrijke beperkende factor in niet-experimenteel onderzoek naar oorzaak en gevolg. Je zou kunnen zeggen dat het resultaat van een eenvoudige vergelijking van de twee groepen de combinatie is van het werkelijke effect van het ene geneesmiddel ten opzichte van het andere én allerhande overwegingen om een diureticum dan wel een ACE-remmer voor te schrijven.

We kunnen dit probleem voorkómen door groepen met elkaar te vergelijken die volkomen vergelijkbaar zijn, behalve ten aanzien van één aspect, bijvoorbeeld de blootstelling aan een geneesmiddel.<sup>39</sup> Als we vervolgens een verschil zien in gezondheidsuitkomsten tussen beide groepen, dan is er maar een manier om dat te verklaren: dat ene aspect waarop de groepen verschillen moet de oorzaak zijn van dat gevonden verschil. Een gerandomiseerd onderzoek, waarbij blootstelling aan bijvoorbeeld middel A of middel B door het lot wordt toegewezen is een manier om die vergelijkbaarheid – gemiddeld genomen – te bereiken.

Een alternatief kan zijn om expliciet rekening te houden met de verschillen tussen groepen mensen die verschillend worden behandeld, bijvoorbeeld in de statistisch analyse van een onderzoek. Om het werkelijke effect van een behandeling te onderscheiden van *confounding* effecten moeten we weten

welke overwegingen tot een bepaalde behandelkeuze hebben geleid. Bij de keuze tussen twee geneesmiddelen spelen mogelijk andere overwegingen dan bij de keuze tussen twee chirurgische technieken. Waar de keuze bij het voorschrijven van een bepaald geneesmiddel voornamelijk wordt gestuurd door patiëntkarakteristieken (in het bijzonder hoe ernstig ziek een patiënt is en in welke mate hij reageerde op eerdere behandelingen), daar gelden andere overwegingen bij de keuze om bijvoorbeeld wel of niet een achillespeesruptuur te opereren, bijvoorbeeld ervaring van de operateur met een van beide technieken. Een ander voorbeeld uit de chirurgische hoek is de behandeling van een sleutelbeenfractuur, waar je operatief een pin in kunt zetten of een plaatje op kunt schroeven om de botdelen bij elkaar te houden. In observationeel onderzoek naar de effecten van deze twee behandelopties lijken de twee behandelgroepen volkomen vergelijkbaar, of een patiënt een pin of een plaat krijgt, lijkt afhankelijk van de voorkeur van de behandelend chirurg. Omdat het min of meer toevallig is door welke dienstdoende chirurg een patiënt wordt behandeld lijkt het probleem van confounding hier veel kleiner. Samen met dr. Marijn Houwert van de afdeling traumatologie van het UMC Utrecht, hebben we dit fenomeen gevonden voor uiteenlopende chirurgische behandelopties in de traumatologie.<sup>40,41</sup>

Nu nog even over die patiënt waarbij het meten van het cholesterolgehalte op zich al een voorspellende waarde lijkt te hebben. We zouden de indruk kunnen krijgen dat het het beste is het cholesterol niet te meten. Dan heeft de patiënt immers de beste prognose. Kan de keuze om een meting te doen de prognose van een patiënt beïnvloeden, ongeacht wat de uitkomst van die meting is? Kan ik door wel of niet te meten de prognose (en dus de toestand) van die individuele patiënt beïnvloeden? Wees gerust het antwoord is ‘nee’. Immers, de groep patiënten waarbij ik wél het cholesterol meet verschilt van de groep patiënten waarbij ik het cholesterol niet meet. De groepen verschillen niet alleen ten aanzien van het wel of niet meten, maar ook ten aanzien van, noem eens wat, lichaamsgewicht, en gemiddelde leeftijd. Dus als we het meten zien als een

oorzaak voor het veranderen van die prognose dan ontstaat een levensgroot risico op confounding.

Kan ik, zonder er een causale duiding aan te geven, stellen dat het niet-meten een voorspellende waarde heeft? Ja, als de overwegingen en redenen om wel of niet te meten dezelfde zijn nu als ten tijde dat het model werd ontwikkeld, dan is er geen probleem en is het niet-meten van bijvoorbeeld het cholesterolgehalte nu net zo informatief als het ooit was. Maar op het moment dat ik dat weet en die kennis beïnvloedt mijn overweging om wel of niet te gaan meten, dan zitten we niet meer in dezelfde situatie. Dit heeft invloed op de zogenaamde transporteerbaarheid van het voorspelmodel. Een voorbeeld hiervan is Google Flu.<sup>42</sup> Op basis van zoekgedrag via zijn zoekmachine was Google heel goed in staat nauwkeurig te voorspellen wanneer de volgende griep epidemie in de Verenigde Staten zou uitbreken. Het model, ontwikkeld begin jaren 2000 was voor de rest van dat decennium accuraat, maar daarna werd het allengs minder. De oorzaak hiervoor wordt gezocht in het veranderde en intensiever gebruik van zoekmachines als Google, die we tegenwoordig eenvoudig en overal op onze slimme telefoons kunnen raadplegen. Nu weer terug naar die patiënt waarvan we wel of niet het cholesterol meten. Ook in die situatie zal een verandering in de overwegingen om wel of niet te meten betekenen dat de voorspellende waarde van niet-meten verandert.

Nu hoor ik u bijna denken: “Kom op zeg, wie doet dit nou? Het is toch evident dat het wel of niet meten van het cholesterol niet zo veel invloed kan hebben.” Uw kritische reactie stelt me dan enigszins gerust. Toch kom ik dergelijke voorbeelden regelmatig tegen in mijn werk, maar let op, nooit zo duidelijk. Vaak is het dan verstoep in machine learning abracadabra en aanverwante analysetechnieken. Begrijp me niet verkeerd. In potentie zijn dat zeer waardevolle technieken, maar hun complexiteit in combinatie met de omvang van de huidige databestanden maakt dat we vaak naar een soort *black box* zitten te kijken, waar eerdergenoemde problemen in verstoep zitten.

### *Van problemen naar uitdagingen*

We hebben drie problemen besproken die kunnen optreden wanneer we routinematig verzamelde zorggegevens gebruiken voor epidemiologisch onderzoek: 1.) ontbrekende waarnemingen, 2.) meetfouten en 3.) confounding. Elk van deze problemen kan leiden tot een systematische vertekening of bias. Dergelijke databestanden kennen vrijwel altijd een zekere mate van deze problemen. Daar de ogen voor sluiten en de data analyseren alsof er niets aan de hand is, is mijns inziens oneigenlijk gebruik van dergelijke data.

En geen van de drie problemen wordt opgelost door grotere databestanden.<sup>43</sup> Hoe meer data, des te nauwkeuriger wordt de schatting van bijvoorbeeld het effect van een behandeling. Maar een schatting die systematisch vertekend is, blijft systematisch vertekend, hoe nauwkeurig die schatting ook is. Ik wil er – hopelijk ten overvloede – op wijzen dat de problemen niet ‘uitmiddelen’ als een databestand groter wordt.<sup>44</sup> Een voorbeeld: als we een groep Nederlanders vragen naar hun gewicht (wat stelselmatig te laag wordt gerapporteerd), dan zullen we het gemiddelde gewicht van Nederlandse volwassenen te laag inschatten, ongeacht of er 100 of 100 000 mensen deelnemen aan het onderzoek. Het is daarom voor te stellen dat het gebruik van kwalitatief hoogwaardige informatie over 100 personen (waarbij ze volgens een vast protocol op de weegschaal worden gezet) de voorkeur heeft boven het gebruik van informatie van beperkte kwaliteit over 100 000 personen.<sup>45</sup>

We kunnen de drie problemen – ontbrekende waarnemingen, meetfouten en confounding – ook zien als uitdagingen waarvoor oplossingen bestaan, die er grofweg op neerkomen dat we externe kennis of extra informatie betrekken in de statistische analyse. Dat is bijvoorbeeld informatie over patiëntkarakteristieken (denk aan het lichaamsgewicht en hoe vitaal een patiënt oogt), maar ook meer ongrijpbare zaken als de interactie tussen de patiënt en zijn behandelaar en de impliciete keuzes die worden gemaakt in de spreekkamer. De bestaande statistische methoden om daar rekening mee te houden zijn ontwikkeld

en getest in een context waarin slechts één van de drie problemen speelt. Toch zullen we in de praktijk vaak worden geconfronteerd met alle drie de problemen tegelijkertijd. Dé vraag wordt dan welke methoden te gebruiken, in welke volgorde ze moeten worden ingezet en hoe verschillende methoden te combineren.

In plaats van nieuwe methoden te ontwikkelen, zouden we energie moeten stoppen in een plaatsbepaling van de reeds bestaande methoden. Een voorbeeld daarvan is het onderzoek dat ik met Bas Penning de Vries deed naar oplossingen voor de combinatie van confounding en ontbrekende waarnemingen,<sup>46</sup> wat de opmaat was voor een door ZonMW gefinancierd Vidi-project. In het kader van dat project werken dr. Maarten van Smeden en promovendi Bas Penning de Vries, Linda Nab en Kim Luijken samen met mij aan juist dit probleem: hoe kunnen we bestaande statistische methoden slim combineren als we worden geconfronteerd met de combinatie van meetfouten, ontbrekende waarnemingen en confounding? Met dit onderzoek hoop ik het risico op oneigenlijk gebruik van grote databestanden van routinematig verkregen zorggegevens te verkleinen en daarmee het nut van de werkzame stof te vergroten. Helaas ontbreekt het mij op dit moment aan de tijd om de eerste resultaten van dit onderzoek hier voor u te schetsen.<sup>47-51</sup> Ik hoop daar op een later moment bij stil te kunnen staan.

### **Overdosering en contra-indicaties**

Terug naar de bijsluiter. Ik wil kort stil staan bij overdosering en contra-indicaties. Zoals ik al eerder in deze rede aangaf: een dataset is tegenwoordig zo gevonden en statistische software is gratis beschikbaar op het internet. Hoewel dit veel mogelijkheden schept, draagt het ook een risico in zich: het risico van *maar eens wat proberen* en de resultaten vervolgens presenteren als de bevestiging van een duidelijke hypothese. Zoek lang genoeg en je vindt vanzelf een aansprekend onderzoeksresultaat. Ik noem dit overdosering: het gebruiken van data voor te veel doeleinden, dat wil zeggen het te vaak gebruiken van data. Het gevolg is dat de wetenschappelijke literatuur wordt over-

spoeld door onderzoeksresultaten die accuraat lijken (door de grote mate van precisie die weer het gevolg is van de omvang van de datasets), maar waarschijnlijk onderhevig zijn aan de eerdergenoemde vertekening door de trias van ontbrekende waarnemingen, meetfouten en confounding. Toch kan een dergelijke aanpak waardevol zijn als de toevalstreffers een werkelijk signaal blijken te zijn bij veelvuldige replicatie. Denk bijvoorbeeld aan genetische oorzaken van ziekten. De uitdaging zit er in om als onderzoeker transparant te zijn: is er sprake van een bevestiging van een vooraf opgestelde hypothese of eigenlijk van een zoektocht naar een nieuwe hypothese?

Voor elk geneesmiddel zijn er situaties waarin het middel juist niet moet worden gebruikt, de contra-indicaties. Net zo zijn er situaties voor te stellen waarin we onderzoeksgegevens juist niet moeten gebruiken om bepaalde onderzoeksvragen te beantwoorden. Ik geef u een paar voorbeelden die ik daarvan heb gezien. Ik wil onderzoek doen naar de effecten van bloedverdunners, maar de reden voor het gebruik van bloedverdunners ken ik niet. Of ik wil onderzoek doen naar de relatie tussen persoonlijkheidskenmerken en gezondheidsklachten en slechts van 10% van mijn onderzoeksgroep ken ik het persoonlijkheidskenmerk. Of ik wil onderzoeken wat belangrijke voorspellers zijn voor de tijd die patiënten doorbrengen op de spoedeisende hulp, maar over de verblijfsduur op de spoedeisende hulp weet ik slechts of die korter of langer was dan 4 uur. Allemaal situaties waarin ik waarschijnlijk zal moeten concluderen dat de beschikbare onderzoeksgegevens van onvoldoende kwaliteit zijn. Ik zal af moeten zien van het onderzoek of – *old school* – zelf het veld in gaan om aanvullende gegevens te verzamelen.

### **Bijwerkingen**

Dat brengt mij bij het derde deel van deze rede, waarin ik wil spreken over mogelijke bijwerkingen van het gebruik van grote databestanden met routinematig verzamelde zorggegevens. Ondeugdelijk onderzoek is potentieel gevaarlijk. Een speciale waarschuwing is dan ook op zijn plaats. Een populair onder-

werp in het publieke debat is vertrouwen in de wetenschap, of wellicht beter gezegd vertrouwen in het wetenschappelijk onderzoekstelsel, wetenschappelijke onderzoeksinstituten, of nog concreter, wetenschappers. We moeten voorzichtig met dat vertrouwen omgaan, want zoals u weet: vertrouwen komt te voet en gaat te paard.

Bij de term ‘niet-integer wetenschappelijk gedrag’ wordt al snel gedacht aan het verzinnen van onderzoeksresultaten en dat presenteren als de opbrengst van gedegen onderzoek. Zeer onfatsoenlijk, maar een relatief voordeel is dat dit een bewuste actie vereist, waardoor er een drempel ontstaat. Ik denk dan ook dat dit in de praktijk zeer uitzonderlijk is. Een probleem dat mijns inziens veel meer impact heeft is het onjuist gebruik van methoden of het gebruik van onjuiste methoden of onjuiste data om een bepaalde onderzoeksvraag te beantwoorden.<sup>52</sup> Tegelijkertijd wordt dit niet bewust gedaan, u mag het onbewust onbekwaam noemen.

Oplossingen voor onbewuste onbekwaamheid liggen niet voor het oprapen. Deels liggen ze in het werken met onderzoeksprotocollen, in transparantie bij opzet, uitvoering en rapportage van onderzoek, maar bovenal in samenwerking. Middels protocollen kan voorafgaand aan een onderzoek een methodologische check worden ingebouwd. Bij transparantie denk ik onder meer aan het beschikbaar stellen van analysecode en uittreksels van databestanden, waardoor het in ieder geval voor onderzoekers die niet betrokken waren bij de uitvoering van het onderzoek mogelijk is een waardeoordeel te vellen over dat onderzoek. Maar het meest hecht ik aan samenwerking tussen disciplines: als ik een voorspelling wil maken van het cardiovasculaire risico bij een patiënt die zich meldt op het spreekuur van een huisarts, dan is dat risico niet een eenvoudige optelsom van datapunten. Het is óók een afspiegeling van meer ongrijpbare gebeurtenissen en impliciete keuzes die door huisarts én patiënt worden gemaakt in de spreekkamer. Om ook daar rekening mee te kunnen houden en zodoende werkelijk relevante informatie uit medische zorgdata te krijgen is sa-

menwerking tussen disciplines onontbeerlijk, in dit voorbeeld een samenwerking tussen ten minste huisartsonderzoekers en methodologen. Sterker nog, beiden hebben elkaar nodig. Methodologen moeten niet aan de haal gaan met beschikbare datasets en huisartsen moeten niet op eigen houtje methodologisch ingewikkeld onderzoek doen.

Ik wil mij de komende jaren inzetten om methoden te ontwikkelen die tegelijkertijd de trias van ontbrekende waarnemingen, meetfouten en confounding aanpakken, zodat we relevante informatie kunnen halen uit routinematig verzamelde zorggegevens. Daarnaast wil ik mij inzetten voor samenwerkingen met onderzoekers uit de nulde, eerste en tweede lijn om de effecten van medische behandelingen te evalueren. Dit ga ik doen door bestaande samenwerkingen voort te zetten en nieuwe samenwerkingen aan te gaan om nog beter te snappen hoe data ontstaan in de spreekkamer en wat voor effect dat heeft op de dataset als geheel. En ik wil mij inspannen voor het opleiden van nieuwe cohorten biomedisch onderzoekers, want zij zijn het die in toenemende mate toegang zullen hebben tot omvangrijke databestanden en geavanceerde data-analysetechnieken. Deze zijn geen panacee en dienen te worden voorzien van een bijsluiters zoals ik zojuist heb betoogd.

### **Dankwoord**

Ik wil eindigen met enkele woorden van dank.

Het College van Bestuur van de Universiteit Leiden en de Raad van Bestuur van het LUMC dank ik voor het in mij gestelde vertrouwen en de hartelijke ontvangst in Leiden.

Mijn promotor de hooggeleerde Arno Hoes en mijn copromotor de hooggeleerde Eelko Hak wil ik bedanken voor het met zorg begeleiden van mijn eerste passen op het pad der epidemiologie.

De hooggeleerde Carl Moons en de hooggeleerde Olaf Klungel wil ik bedanken voor de jarenlange samenwerking en inspiratie in de jaren na mijn promotie. I would like to thank professor Kate Tilling for our fun and fruitful collaborations in the past years.

De hooggeleerde Frits Rosendaal dank ik voor de gastvrije ontvangst op de afdeling Klinische Epidemiologie en zijn snelle besluitvorming. Ik hoop dat we in de toekomst nog veel epidemiologische discussies zullen voeren. De hooggeleerde Jan Vandenbroucke en de hooggeleerde Ewout Steyerberg wil ik bedanken voor hun adviezen, telefonisch of in de Rocky Mountains.

Ook dank aan al mijn oud-collegae bij het Julius Centrum van het UMC Utrecht, en mijn nieuwe collegae bij het LUMC, de afdeling Klinische Epidemiologie en de afdeling Biomedical Data Sciences. Zonder jullie zou het maar een saaie boel worden. In het bijzonder dank ik de hooggeleerde Yolanda van der Graaf, de zeergeleerde Linda Peelen en de zeergeleerde Lotty Hoofst voor een luisterend oor en eerlijke feedback.

Dank aan alle promovendi die ik de afgelopen jaren heb mogen begeleiden en van wie en met wie ik ontzettend veel heb mogen leren.

Yvonne Souverein, dank voor de hulp bij de organisatie van vandaag.

Bas Penning de Vries, Linda Nab, Kim Luijken en de zeergeleerde Maarten van Smeden wil ik bedanken voor hun loyaliteit en hun bereidheid om samen met mij het Leidse avontuur aan te gaan.

Tot slot, wil ik mijn vrienden en familie bedanken dat ze er vandaag zijn. In het bijzonder wil ik bedanken de zeergeleerde Joé Kolkert en de zeergeleerde Elske van Gils. Jullie jarenlange steun, al komt die vanaf de andere kant van de wereld, waardeer ik zeer. Mijn schoonouders wil ik bedanken voor hun leuke dochter en hun liefdevolle hulp bij de zorg voor onze kinderen. Mijn zus dank ik voor haar redactionele commentaar op deze rede. Mijn ouders wil ik bedanken voor de vrijheid om dromen na te jagen en hun onvoorwaardelijke support als dat soms lastig blijkt. Janneke en Floris, dank jullie wel voor al jullie gezelligheid, grappen en liefde. Marieke, van de Grote Markt naar het Rapenburg, de tijd vliegt voorbij maar de liefde blijft. Zonder jou had ik hier niet gestaan, dank je wel.

Allen hier aanwezig: ik dank u voor uw aandacht.

## Referenties

1. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. *Health Inf Sci Syst.* 2014;2:3.
2. Dinov ID. Volume and Value of Big Healthcare Data. *J Med Stat Inform.* 2016;4.
3. Groenwold RH, Hoes AW, Nichol KL, Hak E. Quantifying the potential role of unmeasured confounders: the example of influenza vaccination. *Int J Epidemiol.* 2008;37(6):1422-9.
4. Ludvigsson JF, Adami HO. The urgency to embrace Big Data opportunities in medicine. *J Intern Med.* 2018;283(5):479-80.
5. Beam AL, Kohane IS. Big Data and Machine Learning in Health Care. *JAMA.* 2018;319(13):1317-8.
6. McKinstry B. All watched over by machines of loving grace: an optimistic view of big data. *Br Med J.* 2017;358:j3967.
7. Groenwold RHH, Nab L, van Smeden M. Groot, groter, grootst: big data in medisch onderzoek. *Ned Tijdschr Geneesk.* 2018;162:D3108.
8. Altman DG, Bland JM. Uncertainty and sampling error. *Br Med J.* 2014;349:g7064.
9. Vandembroucke JP, Hofman A, Stiphout WAHJ. *Grondslagen der epidemiologie.* Elsevier gezondheidszorg, 2006.
10. Grobbee DE, Hoes AW. *Clinical epidemiology: principles, methods, and applications for clinical research.* Jones & Bartlett Learning, 2009.
11. *Clinical Practice Research Datalink.* <https://www.cprd.com/> [bezocht op 25 januari 2019].
12. Klungel OH, Kurz X, de Groot MC, Schlienger RG, Tcherny-Lessenot S, Grimaldi L, Ibáñez L, Groenwold RH, Reynolds RF. Multi-centre, multi-database studies with common protocols: lessons learnt from the IMI PROTECT project. *Pharmacoepidemiol Drug Saf.* 2016;25 Suppl 1:156-65.
13. Uddin MJ, Groenwold RH, de Boer A, Gardarsdottir H, Martin E, Candore G, Belitser SV, Hoes AW, Roes KC, Klungel OH. Instrumental variables analysis using multiple databases: an example of antidepressant use and risk of hip fracture. *Pharmacoepidemiol Drug Saf.* 2016;25 Suppl 1:122-31.
14. Uddin MJ, Groenwold RH, de Boer A, Afonso AS, Primates P, Becker C, Belitser SV, Hoes AW, Roes KC, Klungel OH. Evaluating different physician's prescribing preference based instrumental variables in two primary care databases: a study of inhaled long-acting beta2-agonist use and the risk of myocardial infarction. *Pharmacoepidemiol Drug Saf.* 2016;25 Suppl 1:132-41.
15. Haagse Vaten. <http://www.haagsevaten.nl> [bezocht op 17 januari 2019].
16. Flinterman L, Gaag M van der, Opstelten W, Nielen M, Korevaar J. *Kleine kwalen in de huisartsenpraktijk: contacten en medicatiebeleid voor de jaren 2011, 2013 en 2015.* Utrecht: NIVEL, 2017.
17. Keiding N, Louis TA. Perils and potentials of self-selected entry to epidemiological studies and surveys. *J Roy Stat Soc A.* 2016;179:319-76.
18. Hemingway H, Asselbergs FW, Danesh J, Dobson R, Maniadakis N, Maggioni A, van Thiel GJM, Cronin M, Brobert G, Vardas P, Anker SD, Grobbee DE, Denaxas S; Innovative Medicines Initiative 2nd programme, Big Data for Better Outcomes, BigData@Heart Consortium of 20 academic and industry partners including ESC. Big data from electronic health records for early and late translational cardiovascular research: challenges and potential. *Eur Heart J.* 2018;39(16):1481-95.
19. Wallert J, Tomasoni M, Madison G, Held C. Predicting two-year survival versus non-survival after first myocardial infarction using machine learning and Swedish national register data. *BMC Med Inform Decis Mak.* 2017;17(1):99.
20. Groenwold RHH. Drie vormen van bias. *Ned Tijdschr Geneesk.* 2013;157:A6497.
21. NTvG. Wat zijn missende waarden? <https://www.youtube.com/watch?v=XscA0g1my7o> [bezocht op 25 januari 2019].

22. Little RJA, Rubin DB. Statistical analysis with missing data. John Wiley & Sons, 2014.
23. Donders AR, van der Heijden GJ, Stijnen T, Moons KG. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol.* 2006;59(10):1087-91.
24. Brakenhoff TB, Mitroiu M, Keogh RH, Moons KGM, Groenwold RHH, van Smeden M. Measurement error is often neglected in medical literature: a systematic review. *J Clin Epidemiol.* 2018;98:89-97.
25. May AM, Barnes DR, Forouhi NG, Luben R, Khaw KT, Wareham NJ, Peeters PH, Sharp SJ. Prediction of measured weight from self-reported weight was not improved after stratification by body mass index. *Obesity (Silver Spring).* 2013;21(1):E137-42.
26. Carroll RJ, Ruppert D, Crainiceanu CM, Stefanski LA. Measurement error in nonlinear models: a modern perspective. Chapman and Hall/CRC, 2006.
27. Brakenhoff TB, van Smeden M, Visseren FLJ, Groenwold RHH. Random measurement error: Why worry? An example of cardiovascular risk factors. *PLoS One.* 2018;13(2):e0192298.
28. Marcum ZA, Sevick MA, Handler SM. Medication nonadherence: a diagnosable and treatable medical condition. *JAMA.* 2013;309(20):2105-6.
29. Osterberg L, Blaschke T. Adherence to medication. *N Engl J Med.* 2005;353(5):487-97.
30. Lévesque LE, Hanley JA, Kezouh A, Suissa S. Problem of immortal time bias in cohort studies: example using statins for preventing progression of diabetes. *Br Med J.* 2010;340:b5087.
31. Agniel D, Kohane IS, Weber GM. Biases in electronic health record data due to processes within the health-care system: retrospective observational study. *Br Med J.* 2018;361:k1479.
32. Nederlands Huisartsen Genootschap. Cardiovasculair risicomanagement, in herziening 2012. <https://www.nhg.org/standaarden/samenvatting/cardiovasculair-risicomanagement> [bezocht op 17 januari 2019].
33. Groenwold RH, Klungel OH, Altman DG, van der Graaf Y, Hoes AW, Moons KG; PROTECT WP2 (Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium, Work Programme 2 [Framework for pharmacoepidemiology studies]). Adjustment for continuous confounders: an example of how to prevent residual confounding. *CMAJ.* 2013;185(5):401-6.
34. Groenwold RH, Hoes AW, Hak E. Impact of influenza vaccination on mortality risk among the elderly. *Eur Respir J.* 2009;34(1):56-62.
35. Penning de Vries BBL, Kolkert JLP, Meerwaldt R, Groenwold RHH. Atmospheric Pressure and Abdominal Aortic Aneurysm Rupture: Results From a Time Series Analysis and Case-Crossover Study. *Vasc Endovascular Surg.* 2017;51(7):441-6.
36. Groenwold RHH. Verstoring in observationeel onderzoek: 'confounding'. *Ned Tijdschr Geneeskd.* 2012;156:A4221.
37. Grobbee DE, Hoes AW. Confounding and indication for treatment in evaluation of drug treatment for hypertension. *Br Med J.* 1997;315(7116):1151-4.
38. Vandembroucke JP. When are observational studies as credible as randomised trials? *Lancet.* 2004;363(9422):1728-31.
39. Hernán MA, Robins JM. Causal Inference. Boca Raton: Chapman & Hall/CRC, 2019.
40. Ochen Y, Beks RB, van Heijl M, Hietbrink F, Leenen LPH, van der Velde D, Heng M, van der Meijden O, Groenwold RHH, Houwert RM. Operative treatment versus nonoperative treatment of Achilles tendon ruptures: systematic review and meta-analysis. *Br Med J.* 2019;364:k5120.
41. Beks RB, Houwert RM, Groenwold RHH. Meerwaarde van observationeel onderzoek in chirurgie. *Ned Tijdschr Geneeskd.* 2017;161:D1493.
42. Lazer D, Kennedy R, King G, Vespignani A. The Parable of Google Flu: Traps in Big Data Analysis. *Science.* 2014;343(6176): 1203-5.
43. Altman DG, Bland JM. Uncertainty beyond sampling error. *Br Med J.* 2014;349:g7065.



44. Loken E, Gelman A. Measurement error and the replication crisis. *Science*. 2017;355(6325):584-5.
45. Shah ND, Steyerberg EW, Kent DM. Big Data and Predictive Analytics: Recalibrating Expectations. *JAMA*. 2018;320(1):27-8.
46. Penning de Vries B, Groenwold R. Comments on propensity score matching following multiple imputation. *Stat Methods Med Res*. 2016;25(6):3066-8.
47. Luijken K, Groenwold RH, van Calster B, Steyerberg EW, van Smeden M. Impact of predictor measurement heterogeneity across settings on performance of prediction models: a measurement error perspective. *arXiv preprint arXiv:1806.10495*, 2018.
48. Nab L, Groenwold RH, Welsing PM, van Smeden M. Measurement error in continuous endpoints in randomised trials: problems and solutions. *arXiv preprint arXiv:1809.07068*, 2018.
49. Penning de Vries BBL, van Smeden M, Groenwold RHH. A weighting method for simultaneous adjustment for confounding and joint exposure-outcome misclassifications. *arXiv preprint arXiv:1901.04795*, 2019.
50. de Vries BBP, van Smeden M, Groenwold RH. Propensity score estimation using classification and regression trees in the presence of missing covariate data. *Epidemiologic Methods*. 2018.
51. Groenwold RH, Shofty I, Miočević M, van Smeden M, Klugkist I. Adjustment for unmeasured confounding through informative priors for the confounder-outcome relation. *BMC Med Res Method*. 2018;18(1), 174.
52. Altman DG. Statistics and ethics in medical research. Misuse of statistics is unethical. *Br Med J*. 1980;281(6249):1182-4.





## PROF.DR. R.H.H. GROENWOLD



- 2004 BSc Natuurkunde, Rijksuniversiteit Groningen
- 2004 Artsexamen, Rijksuniversiteit Groningen
- 2008 MSc Epidemiologie, Universiteit Utrecht
- 2009 Promotie faculteit Geneeskunde, Universiteit Utrecht
- 2015 MSc Statistics, University of Sheffield, UK
- 2009 - 2017 universitair docent en universitair hoofddocent, Julius Centrum voor Gezondheidswetenschappen en Eerstelijns Geneeskunde, UMC Utrecht.
- 2017 - 2018 universitair hoofddocent, afdeling Klinische Epidemiologie, LUMC Leiden
- 2018 – heden Hoogleraar Klinische Epidemiologie, in het bijzonder methodologie van onderzoek naar medische behandelingen, LUMC Leiden



Universiteit  
Leiden