

# Het individu ontcijferd



Rede uitgesproken door

Prof. dr. Ellen L. Hamaker

bij de aanvaarding van het ambt van hoogleraar op het gebied van

*Longitudinal Data Analysis*

in de Faculteit Sociale Wetenschappen

aan de Universiteit Utrecht

op vrijdag 12 april 2019.



Universiteit Utrecht

Mijnheer de Rector Magnificus,

mijnheer de decaan,

geachte aanwezigen,

“Men vergelijkt het arsenaal van statistische methoden wel met een kist met gereedschap. Die vergelijking is leerzaam. Het gaat erom bij een gegeven probleem het juiste gereedschap te kiezen en dit op de juiste wijze te hanteren.” (H. C. Hamaker, 1961).

Dit is een citaat uit de oratie van mijn grootvader, Hugo Christiaan Hamaker, die hij uitsprak toen hij in 1960 hoogleraar *Toegepaste Statistiek* werd.

“Aha,” zullen sommigen van u nu denken, “het zit duidelijk in haar genen!” Maar anderen zullen deze conclusie voorbarig vinden. Zij zullen tegenwerpen dat mijn liefde voor cijfers en formules niet per se genetisch bepaald is, maar me ook heel goed met de paplepel kan zijn ingegoten. De vraag of menselijke eigenschappen zoals intelligentie of persoonlijkheid erfelijk zijn, óf dat zij het resultaat zijn van opvoeding en andere omgevingsinvloeden, wordt ook wel het nature-nurture vraagstuk genoemd. En het was precies dit vraagstuk wat meer dan een eeuw geleden de aanzet vormde voor het ontwikkelen van de statistische methoden zoals wij die vandaag de dag nog steeds gebruiken en onderwijzen in de sociale wetenschappen.

Eén van de belangrijkste grondleggers van de hedendaagse statistiek was Sir Francis Galton. Hij was sterk geïnspireerd door de evolutietheorie van zijn neef Charles Darwin, en wilde het concept van natuurlijke selectie toepassen op menselijke eigenschappen (Cowles, 1989; Desrosières, 1998). Galton was ervan overtuigd dat evolutionaire “fitness” gelijk staat aan intelligentie. De vraag werd daarom: Hoe kunnen we intelligentie goed meten, en hoe kunnen we de mate van erfelijkheid van intelligentie bepalen?

Galton kreeg hiervoor hulp van Karl Pearson, die in 1895 een maat ontwikkelde waarmee we de samenhang tussen twee variabelen—zoals de intelligentie van vaders en hun zonen—kunnen weergeven in één getal: de correlatiecoëfficiënt. Daarnaast leverde Charles Spearman een belangrijke bijdrage door in 1904 de techniek van factor analyse te presenteren, waarmee de meetfout gescheiden kan worden van de ware onderliggende score. Een vierde belangrijke bijdrage kwam van Sir Ronald Fisher, die zich realiseerde dat steekproeven geen exacte kopieën zijn van de populatie waar zij uitkomen. Hij ontwikkelde daarom inferentiële statistiek, waarmee het mogelijk is om op grond van een steekproef kansuitspraken over de populatie te doen.

Deze *founding fathers* van de statistiek werden alle vier sterk gedreven door een interesse in erfelijkheid en natuurlijke selectie (Cowles, 1989; Desrosières, 1998). Dit is een belangrijk detail, omdat het van grote invloed is geweest op hun werk, en op de statistische methoden die zij hebben ontwikkeld. Natuurlijke selectie is een mechanisme dat opereert op het niveau van de populatie; het gaat er niet zo zeer om dat een specifiek individu volwassen wordt en zich voortplant, maar of bepaalde genen en de daarbij behorende eigenschappen in een populatie behouden blijven of uitsterven. De grondleggers waren hier ook heel duidelijk in: Zij schreven dat het doel was om niet het individu te onderzoeken, maar een soort (Weldon, Pearson & Davenport, 1901). Dit is waar zij statistiek voor ontwikkelden.

In de sociale wetenschappen hebben we allerlei vragen die zich op populatieniveau afspelen. Een goed voorbeeld hiervan zijn *selectievraagstukken*, waarbij we willen bepalen welke mensen geschikt zijn voor een bepaalde baan of opleiding, wie er risico loopt op een depressie of om schoolverlater of extremist te worden, of wie gebaat zal zijn bij behandeling met pillen of juist met praten. Maar dit is slechts een deel van onze vraagstukken.

Veel van onze theorieën gaan juist over het individu, over zijn of haar gedrag, gedachten en gevoelens, vaak in wisselwerking met de omgeving. Een goed voorbeeld hiervan zijn de motivatiesystemen die worden veronderstelt ons gedrag te sturen (Carver & White, 1994). Enerzijds is er het *behavioral inhibition system* wat in actie treedt als we negatieve consequenties verwachten, en wat ervoor zorgt dat we bepaald gedrag niet vertonen om op die manier straf te vermijden; dit zorgt er nu bijvoorbeeld voor dat u braaf op uw stoel blijft zitten en niet uitgebreid gaat zitten bellen of twitteren. Anderzijds is er het *behavioral approach system*, wat in actie treedt als we beloning verwachten; dit zal er straks voor zorgen dat u op staat en in de rij gaat staan bij de receptie, waar u iets te eten en te drinken hoopt te krijgen.

Zo zijn er talloze voorbeelden te geven van theorieën in de sociale wetenschappen die niet over de populatie gaan, maar over het individu. En we hebben dus een kist vol gereedschap geërfd dat niet goed past bij veel van de klussen die we willen klaren. Toch is dit waar we het tot voor kort grotendeels mee hebben moeten doen, met alle gevolgen van dien. Want u kunt zich voorstellen dat als de klus eigenlijk om een waterpomptang vraagt, het niet mooi wordt als u alleen een beitel en een hamer heeft.

## Is dat nou echt zo erg?

Om hier antwoord op te kunnen geven, is het goed om onze gebruikelijke methoden van onderzoek van wat dichterbij te bekijken. Er zijn grofweg drie belangrijke onderzoeksmethodes die we kunnen onderscheiden.

Ten eerste is er *cross-sectioneel onderzoek*. Dit vormt de basis van veelgebruikte statistische technieken, zoals regressie analyse en factor analyse. Het is gebaseerd op het nemen van een steekproef van personen, om vervolgens bepaalde kenmerken bij die personen te meten. Een bekend voorbeeld dat ik hierbij vaak aanhaal, is het typevoorbeeld (Hamaker, 2012). Stel dat we een groep mensen hebben geselecteerd, en we laten hen allemaal vijf minuten typen; na afloop kijken we bij ieder van hen hoeveel woorden ze hebben getypt, en wat het percentage fouten is dat zij hebben gemaakt.

Dit is een prima methode als we geïnteresseerd zijn in selectie: we zien dat mensen die sneller typen doorgaans minder fouten maken. Maar als we vervolgens voor een specifiek persoon willen weten hoe we ervoor kunnen zorgen dat deze persoon minder fouten maakt, hebben we bar weinig aan deze resultaten. Want u voelt natuurlijk haarfijn aan dat als iemand sneller probeert te typen, er meer fouten gemaakt gaan worden. Als ons doel is om te *beschrijven* en te *selecteren*, dan is er niets mis met dit soort cross-sectioneel onderzoek; maar als we inzicht willen krijgen in de onderliggende mechanismen, als we willen *begrijpen*, en misschien zelfs willen kunnen *ingrijpen*, dan schiet deze aanpak in de meeste gevallen hopeloos tekort.

Soms werpt men tegen dat het in veel gevallen niet zo dramatisch zal zijn als in het typevoorbeeld, en dat een negatieve samenhang op populatieniveau heel vaak samen zal gaan met een negatieve samenhang binnen het individu. Kijken we bijvoorbeeld naar de relatie tussen stress en geluk in de populatie, dan vinden we dat personen die hoog op stress zitten vaak minder gelukkig zijn; en binnen een individu zullen we waarschijnlijk ook zien dat op stressvolle momenten de persoon minder gelukkig is. Maar hoewel we zo een gelijkenis waarschijnlijk kunnen vinden, is een dergelijk donkerbruin vermoeden natuurlijk nog steeds geen wetenschap. Zolang we er geen onderzoek naar hebben gedaan, is het voorbarig om aan te nemen dat deze relaties wel ongeveer gelijk zullen zijn en dat we daarom op grond van cross-sectioneel onderzoek zinnige uitspraken over processen binnen een individu kunnen doen.

Wat we wél kunnen doen om een proces te onderzoeken, is gebruik maken van een tweede belangrijke onderzoeksmethode: het *experiment*. Hierin worden deelnemers random toegewezen aan een van meerder condities, waarbij het idee is dat, hoewel er allerlei verschillen zijn tussen de deelnemers, deze zich uitmiddelen op groepsniveau. Vervolgens zouden we bijvoorbeeld tegen de

ene groep proefpersonen kunnen zeggen dat zij zo snel mogelijk moeten typen, terwijl we tegen de andere groep niets zeggen. Als we dan na afloop een verschil zien tussen het gemiddelde aantal fouten dat in de beide groepen is gemaakt, kunnen we dat met vrij grote zekerheid toeschrijven aan het verschil in de instructies die we hebben gegeven. Experimenteel onderzoek wordt daarom ook wel gezien als de koninklijke weg om tot causale uitspraken te komen.

Een mooi experiment bedenken is een ambacht waar ik grote bewondering voor heb. Maar toch zijn experimenten niet zaligmakend. Ten eerste is het niet altijd etisch en/of praktisch om een experiment uit te voeren. Als we willen weten wat het effect is van een scheiding op de ontwikkeling van kinderen, is het niet mogelijk om gezinnen random toe te wijzen aan de conditie “scheiden” en de conditie “niet-scheiden”. Dat geldt voor veel aspecten die we interessant vinden als sociale wetenschappers.

Ten tweede is de uitkomst van een experiment niet altijd zo makkelijk te vertalen naar het dagelijks leven. Om hier een voorbeeld te geven: ik heb zelf ooit deelgenomen aan een goktaak in een laboratorium. Mensen die mij een beetje kennen weten dat ik een heel behoudend, uiterst voorzichtig type ben: ik sta regelmaat met mijn fiets voor een rood stoplicht te wachten terwijl de rest voorbijraast. Dus daar in het lab dacht ik: Laat ik eens niet zo flauw en bangig doen, en ik nam alle riskante beslissingen die mogelijk waren (en ik verloor dan ook heel veel punten). Dit is natuurlijk anekdotisch, maar het illustreert wel dat de gecontroleerde omgeving van een lab wellicht ander gedrag uitlokt bij deelnemers dan wat normaal is voor hen, en waar we eigenlijk in geïnteresseerd zijn.

Een derde kanttekening bij het gebruik van experimenten is dat het vooral een beeld geeft van een uitkomst van een proces, zonder dat we echt inzicht krijgen in hoe het proces precies verloopt over de tijd.

Om meer inzicht te krijgen in het verloop van een proces over de tijd, maken sociale wetenschappers gebruik van een derde belangrijke onderzoeksmethode: *panel onderzoek*, ofwel *longitudinaal onderzoek*. Kenmerkend hierbij is dat dezelfde personen meerdere keren worden gemeten. Hiermee kunnen we bijvoorbeeld de ontwikkeling van kinderen volgen door hun groeicurves in kaart te brengen, en bekijken welke factoren voorspellend zijn hiervoor. Deze inzichten kunnen we dan weer gebruiken om te bepalen of specifieke kinderen het risico lopen om af te haken, en om eerder in te kunnen grijpen. Maar om te weten wát we dan zouden moeten doen, hoe we het beste kunnen interveniëren, is het belangrijk dat we begrijpen hoe de onderliggende causale structuur in elkaar steekt.

Voorbeelden van vragen op dit vlak zijn: leidt cannabis gebruik tot een verhoogde kans op psychoses en depressies, of leiden dergelijke psychische stoornissen tot meer cannabis gebruik? Een andere vraag: zorgen strenge opvoedmethodes—zoals schreeuwen, fysieke straffen, en het vernederen van kinderen—tot meer of juist minder probleemgedrag bij kinderen? En zijn ouders van moeilijke kinderen eerder geneigd om dergelijke opvoedmethodes te gebruiken? En nog een andere vraag: Leiden zorgen over de eigen gezondheid tot meer online zoekgedrag naar allerlei symptomen en ziektes, of leiden dergelijke zoektochten juist tot meer zorgen?

Om dit soort vragen te onderzoeken wordt sinds de jaren 80 veel gebruik gemaakt van het cross-lagged panel model (Rogosa, 1980). Maar in 2015 heb ik samen met Rebecca Kuiper en Raoul Grasman een artikel gepubliceerd waarin we dit model bekritisieren en laten zien dat de resultaten die ermee worden verkregen een totaal vertekend beeld kunnen geven van het onderliggende proces: Zo kan er een relatie lijken te bestaan, die er in werkelijkheid niet is, of omgekeerd, je vindt geen relatie terwijl dat die er wel is (Hamaker, Kuiper & Grasman, 2015). Maar je kan ook een positief verband vinden, terwijl het in werkelijkheid een negatief verband is. Dat betekent dus dat als je op grond hiervan suggesties doet voor interventies, dit het tegenovergestelde van het gewenste effect zou kunnen hebben. Dat was natuurlijk behoorlijk slecht nieuws, maar ons goede

nieuws was dat het eigenlijk heel eenvoudig op te lossen is met een alternatief model, dat we het random intercept cross-lagged panel model hebben genoemd.

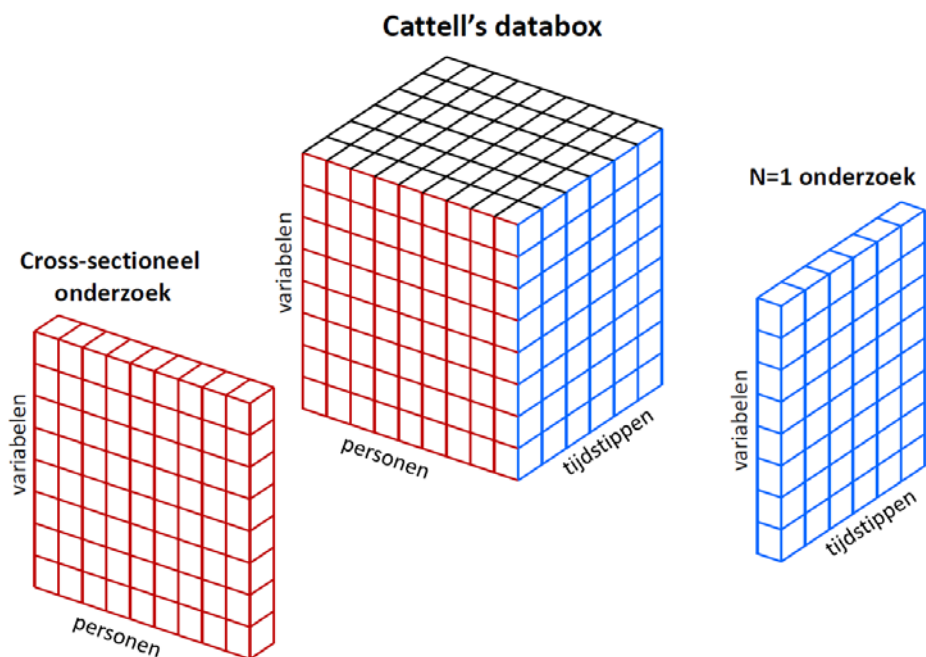
Ik wil hier graag vermelden dat wij niet de enigen waren die dit model bedacht hebben. Dr. Loes Keijsers kwam rond dezelfde tijd met hetzelfde idee (Keijsers, 2015), en later wees professor Hans Ormel mij erop dat hij al in de jaren 90 artikelen publiceerde waarin dit model gebruikt werd (Ormel & Schaufeli, 1991); zo gaat het vaker met goede ideeën!

Hoewel ik enthousiast ben over deze vorm van longitudinaal onderzoek, heeft ook dit soort onderzoek zijn beperkingen: Wanneer we geïnteresseerd zijn in ingewikkelde processen zoals de opvoeding van kinderen, de wisselwerking tussen angst en depressie, of de vraag of social media gebruik leidt tot stress of dat stress leidt tot social media gebruik, dan zijn die paar snap shots waar traditioneel longitudinaal onderzoek op is gebaseerd niet voldoende. Vaak zitten er weken, maanden of soms zelfs jaren tussen de observaties, waardoor het eigenlijk niet mogelijk is om de onderliggende dynamiek van een proces echt in kaart te brengen.

Samenvattend kunnen we dus stellen dat de drie standaard onderzoeksmethodes in de sociale wetenschappen—dat wil zeggen, cross-sectioneel onderzoek, experimenteel onderzoek, en panel onderzoek—stuk voor stuk heel aantrekkelijke kenmerken hebben, maar ook dat ze alle drie verre van optimaal zijn als we eigenlijk willen weten hoe allerlei dagelijkse processen zich ontploegen over de tijd binnen een individu.

## Is dit dan alles wat ze in de sociale wetenschappen hebben?

Gelukkelijk niet. Tegen al deze stromen in zijn er ook steeds onderzoekers geweest die bereid waren van de gebaande paden af te wijken en op zoek te gaan naar iets nieuws, iets wat ons dichterbij het individu kan brengen. Een belangrijke innovator was Raymond Cattell, die als een van de eersten bedacht hoe je statistiek kunt gebruiken om één individu te onderzoeken (Cattell, Cattell & Rhymer, 1947). Hiervoor introduceerde hij het idee van de *databox* (Cattell, 1952), waaruit duidelijk wordt dat er drie dimensies waaruit wij data kunnen sampelen; personen, variabelen, en tijdstippen. Waar we bij cross-sectioneel onderzoek een doorsnede nemen bestaande uit veel mensen op één tijdstip, was Cattell een van de eersten die besloot om 1 persoon op heel veel tijdstippen te meten. Dit wordt ook wel N=1 data genoemd.



Cattell wilde weten of de onderliggende structuur in de dagelijkse fluctuaties van een persoon overeenkomt met de structuur van verschillen tussen personen zoals die al vaker in cross-sectioneel onderzoek was gevonden. Hij liet daarvoor een factor analyse op de N=1 data laten uitvoeren. Tegenwoordig kan iedere bachelor student dit binnen een seconde door de computer laten doen, maar in 1947 betekende dit dat er een assistent 2 maanden lang aan dit probleem zat te rekenen (Luborsky, 1995). Zo een analyse was dus een heel bewuste, weloverwogen keuze.

Deze vorm van factor analyse is bekend geworden als *Cattell's P-techniek*. Mijn promotor, Peter Molenaar, heeft hier in 1985 een uitbreiding op voorgesteld, die bekend staat als *dynamische factor analyse*; in deze aanpak wordt ook rekening gehouden met de volgorde van de observaties, en de mogelijke samenhang van scores over de tijd (Molenaar, 1985). Daarover later meer.

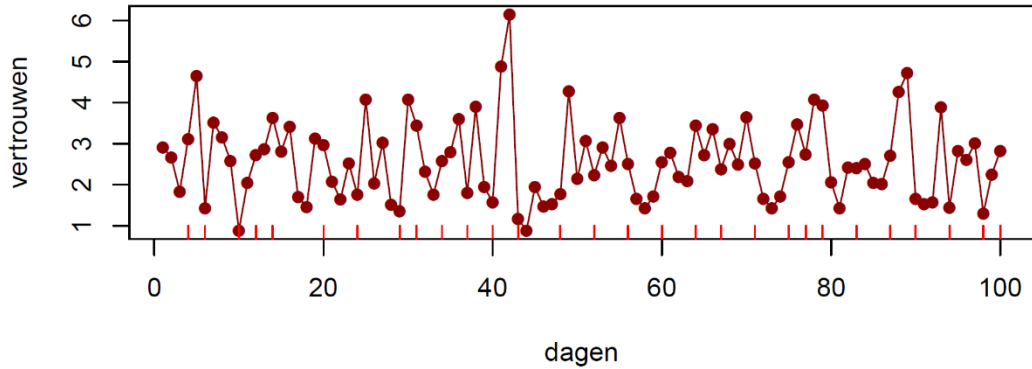
Toch bleef deze N=1 aanpak nog lang exotisch, omdat de data die ervoor nodig zijn, lastig te verzamelen waren. Dat ging toen ongeveer als volgt: Een deelnemer kreeg een stapel vragenlijsten mee naar huis, en een stapel enveloppen met postzegels, en dan moest die persoon iedere dag een vragenlijst invullen en terugsturen naar de onderzoeker. Dit werd zo gedaan om te voorkomen dat de deelnemer op het laatste moment snel de hele stapel invulde. Maar het was dus behoorlijk omslachtig, en er waren nauwelijks onderzoekers die dergelijke data verzamelden. Nog schaarser waren de datasets waarbij dit protocol bij meerdere personen was toegepast, zodat je kon onderzoeken of verschillende personen verschillende patronen hadden.

Dit was ook nog steeds de stand van zaken toen ik in het jaar 2000 begon aan mijn promotieonderzoek onder begeleiding van Peter Molenaar en Conor Dolan. Ik wilde graag de factor structuur en dynamiek van verschillende personen met elkaar vergelijken. Er was één geschikte dataset uit Duitsland waar verschillende onderzoekers gebruik van maakten (Borkenau & Ostendorf, 1998), maar verder was er eigenlijk niets binnen de sociale wetenschappen, waardoor ik op me op den duur genoodzaakt zag om ook maar data van koeien te gaan analyseren. Als ik een presentatie gaf indertijd, kreeg ik stevast na afloop de reactie: Hartstikke leuk werk, maar wie heeft er eigenlijk dit soort data? Ik werkte in feite aan een oplossing voor een probleem dat niet bestond.

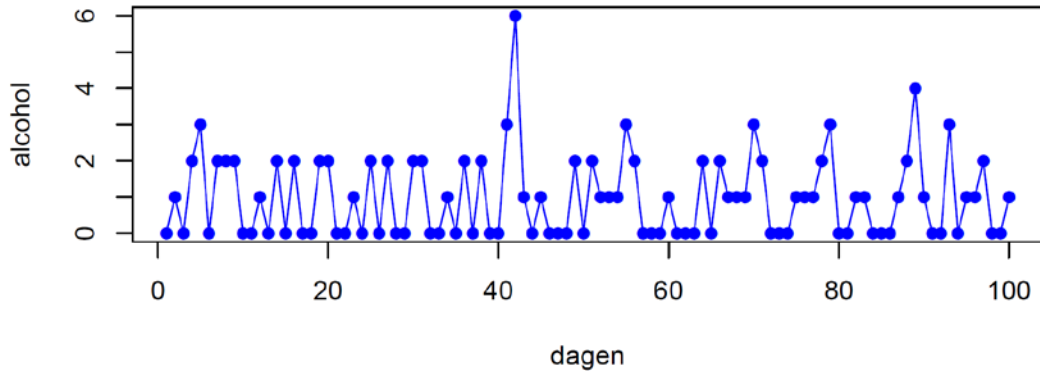
## **Maar dit is sinds mijn promotie 15 jaar geleden totaal veranderd.**

Door smart phones en andere wearable devices zoals stappentellers en hartslagmeters is het mogelijk geworden om mensen gedurende langere tijd te meten in hun natuurlijke habitat, terwijl zij hun gewone dagelijkse leven leiden (Hamaker & Wichers, 2017). Kenmerkend van deze data is dat er heel veel herhaalde metingen per persoon zijn; dit kan variëren van bijvoorbeeld 30 als het gaat om zelf-rapportage, tot honderden of zelfs duizenden metingen als het om bepaalde fysiologische metingen. Deze data worden met allerlei verschillende termen aangeduid, zoals ambulatory assessments of experience sampling; een overkoepelende term is *intensieve longitudinale data* (Collins, 2006). Dergelijke data heeft een aantal aantrekkelijke kenmerken. Ik bespreek er hier twee.

Ten eerste zijn de observaties geordend in de tijd. Hierdoor kunnen we op zoek gaan naar *onderliggende patronen*. Laat me dit illustreren met een concreet voorbeeld: Stel dat ik in de aanloop naar mijn oratie toe dagelijks had gemeten hoeveel vertrouwen ik erin had dat het goed zou komen vandaag. Ik heb dit niet echt gedaan, want ik had eerlijk gezegd al genoeg andere zaken aan mijn hoofd, dus de data die ik hier laat zien heb ik verzonnen. Dat klinkt misschien vreemd en u vraagt zich wellicht af of u nu getuige bent van een geval van fraude, maar in methodologisch onderzoek is het simuleren van data juist een veelgebruikte strategie: Als je namelijk zelf je data maakt, weet je precies wat je erin stopt, en dan kun je onderzoeken welke analyse techniek dat er het beste uit kan halen. En vervolgens adviseren wij andere onderzoekers over. Dit ter toelichting.



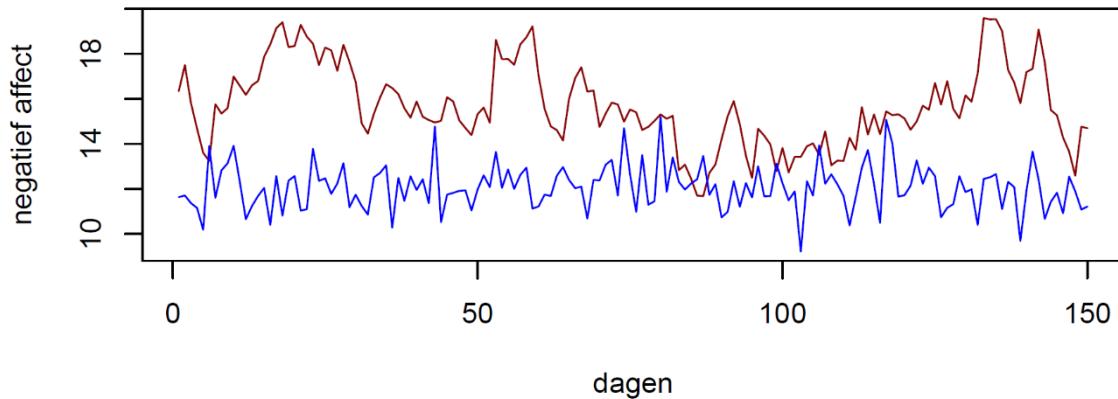
Dus we zien hier de dagelijkse schommelingen in mijn vertrouwen. Nu kunnen we ons afvragen welk patroon hierin schuil gaat, en welke factoren mogelijk van invloed zijn op de fluctuaties. Als u goed kijkt, ziet u onderaan kleine rode streepjes, die aangeven op welke dagen ik gesport heb, en als u dan heel goed kijkt, kunt u misschien zien dat wanneer ik gesport heb, ik vaak de daarop volgende dag wat meer vertrouwen heb in een goede afoop. Maar naast sporten zijn er natuurlijk allerlei andere factoren van invloed.



Hier zien we mijn dagelijkse alcohol consumptie (ik benadruk nogmaals dat dit fictieve data zijn). Het is misschien niet makkelijk met het blote oog te zien, maar met behulp van statistische analyses zouden we vast kunnen stellen dat als ik drink, dit samen gaat met een toename in vertrouwen, maar ook dat hierop volgend er juist een afname in vertrouwen is; een typisch geval van de-rekening-komt-later.

Inzicht in een dergelijk patroon maakt het ook begrijpelijker waarom sommige mensen op moeilijke momenten sneller naar de fles dan naar de sportschoenen grijpen, ook al is het effect daarvan averechts op de langere termijn. En door dit inzichtelijk te maken, zou zo een persoon misschien ook meer gemotiveerd kunnen worden om een regime van regelmatig sporten vol te houden, juist ook op moeilijke momenten. Inzicht in de volgorde waarin gevoelens, gedachten en gedrag zich ontploegen over de tijd kan cruciaal zijn om meer effectieve interventies te ontwikkelen.

Een tweede bijzonder aantrekkelijk aspect van intensieve longitudinale data is dat we *individuele verschillen in de dynamiek* kunnen onderzoeken. Kijk bijvoorbeeld naar de dagelijkse schommelingen in negatieve emoties van deze twee personen. We zien dat deze personen niet alleen van elkaar verschillen qua gemiddelde, en qua variatie, maar we kunnen ook zien dat de ene persoon er langer over doet om weer bij zijn/haar gemiddelde te komen dan de andere.



Deze neiging om ergens in te blijven hangen, wordt ook wel *inertie* genoemd, en kan met behulp van de autoregressie in de data beschreven worden. Professor Peter Kuppens heeft hier veel onderzoek naar gedaan en laten zien dat individuele verschillen in autoregressie in emoties samenhangen met kenmerken als neuroticisme, depressie, en sekse, maar ook dat het voorspellend is voor het ontwikkelen van een depressie in de toekomst, zelfs nadat we hebben gecontroleerd voor het huidige niveau van depressieve klachten (Houben, Van den Noortgate & Kuppens, 2015; Kuppens, Allen & Sheeber, 2010; Kuppens et al., 2012).

Andere onderzoekers hebben gekeken naar individuele verschillen in de mate van *reactiviteit* van mensen. Professor Marieke Wichers heeft bijvoorbeeld gevonden dat mensen die leiden aan een depressie in vergelijking tot gezonder personen sterker op alledaagse negatieve gebeurtenissen reageren, terwijl zij juist minder sterk reageren op positieve alledaagse gebeurtenissen (Geschwind et al., 2010; Wichers et al., 2009; Wichers et al., 2010).

Verder hebben professor Denny Borsboom en dr. Angelique Cramer het idee geopperd dat psychologische stoornissen moeten worden voorgesteld als een *dynamisch netwerk* van symptomen die elkaar over de tijd aansteken (Borsboom, 2017; Borsboom & Cramer, 2013). Het idee hierbij is dat het grote verschil tussen gezond en ziek vooral gezocht moet worden in individuele verschillen in de sterkte van de dynamische connecties tussen de symptomen.

Mijn eigen bijdragen aan dit spannende nieuwe veld zijn tweeledig. Enerzijds houd ik me in mijn werk bezig met de technisch kant, door met behulp van simulaties te bepalen hoe goed modellen geschat kunnen worden en hoe goed ze van elkaar onderscheiden kunnen worden. Hierbij is het ook van belang om te bekijken hoe verschillende modellen analytisch aan elkaar gerelateerd kunnen worden. Ik zal u daar vandaag niet mee vermoeien, maar ik wil wel graag vermelden dat er zijn maar weinig zaken in het leven zo een gelukzaligheid teweeg kunnen brengen als zien hoe de ene formule omgeschreven kan worden in de andere. Anderzijds houd ik me in mijn werk bezig met de conceptuele kant van deze modellen: Wat betekent zo een model nou eigenlijk, en hoe kun je een bepaalde parameter van een model inhoudelijk zinnig interpreteren? Je zou kunnen zeggen dat het eerste vooral de wiskundige in mij is, terwijl het tweede vooral de psycholoog is. Het bouwen van bruggen tussen geavanceerde analyse technieken en inhoudelijke theorie is de belangrijkste rode draad in mijn werk.

Een doorbraak voor mij persoonlijk, maar in mijn ogen ook voor het vakgebied, is gekomen door de ontwikkeling van *dynamic structural equation modeling (DSEM)* in het software programma Mplus. Bengt Muthén en Tihomir Asparouhov hebben mij in 2015 benaderd met de vraag of ik met hen wilde samenwerken aan een nieuwe module om intensieve longitudinale data te kunnen analyseren. Ik heb daar toen wel even over getwijfeld, omdat commerciële software niet goed past in deze tijd van open science. Maar wat voor mij de doorslag gaf, was dat het soort innovaties waar zij toe in staat zijn nu eenmaal veel tijd en expertise vragen, en dat het vrijwel onmogelijk is

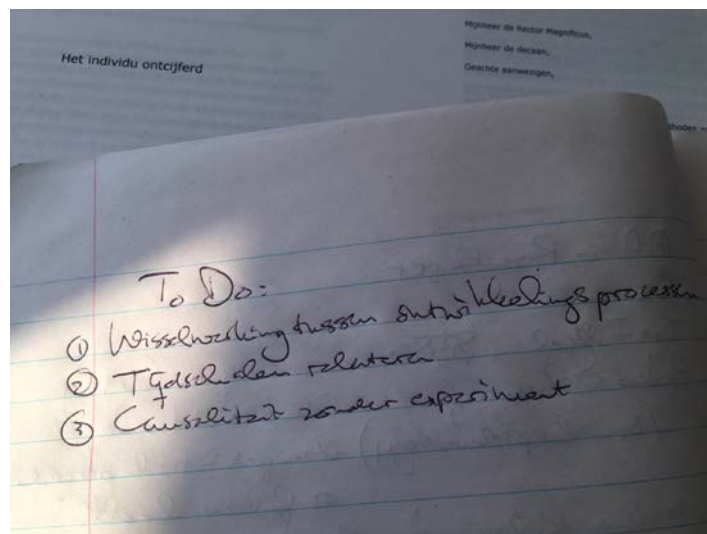


voor wetenschappers om dat “erbij” te doen, naast het gewone onderzoek, het publiceren, het onderwijs, de management taken, en de grote onzekerheid van tijdelijke onderzoekssubsidies.

De samenwerking bleek bijzonder prettig: Zij hadden allerlei vragen, waarvoor ik steeds wel weer een paper wist op te hoesten, niet zelden van mijzelf, omdat ik dezelfde vragen had gehad; en omgekeerd voelde ik me een kind in de snoepwinkel die kon zeggen wat ze allemaal wilde hebben, en bijna altijd haar zin kreeg. Het resultaat is een programma waarmee makkelijk, snel, en accuraat zeer complexe dynamische modellen kunnen worden geschat (Asparouhov, Hamaker & Muthén, 2018). Dit heeft een grote praktische hindernis weggenomen, waardoor het veld van dynamisch modelleren nu voor een breed publiek van onderzoekers toegankelijk is gemaakt.

We zijn een enorm eind gekomen in de 72 jaar sinds Cattell's eerste studie. Het is overduidelijk dat intensief longitudinaal onderzoek veel heeft toe te voegen aan dat wat we al hadden. In vergelijking met cross-sectioneel onderzoek, biedt het de mogelijkheid om stabiele verschillen tussen personen te scheiden van de dynamiek binnen personen. In vergelijking met experimenteel onderzoek, biedt het de mogelijkheid om mensen over langere tijd in hun natuurlijke habitat te onderzoeken. En in vergelijking met het meer traditionele longitudinale onderzoek, biedt het de mogelijkheid om op een veel verfijndere tijdschaal met veel meer observaties per persoon onderzoek te doen naar verschillen tussen personen in hun dynamiek.

## **Wat moet er nog gebeuren, en wat ga ik met mijn leerstoel *Longitudinal Data Analysis* eigenlijk doen?**



Een van de meest basale en tegelijkertijd ook een van de moeilijkste vragen die we in de sociale wetenschappen hebben is hoe we de wederzijdse beïnvloeding tussen twee ontwikkelingsprocessen kunnen onderzoeken. Veel van de bestaande modellen zijn gebaseerd op de aanname dat er geen sprake is van ontwikkeling; het gaat dan om zogenaamde stationaire processen.

Maar dit strookt niet altijd met de werkelijkheid: de leesvaardigheid van een kind neemt toe terwijl ook zijn vermogen om langere tijd stil te zitten in de klas toeneemt, en we willen weten of dat laatste noodzakelijk is voor het eerste; of we zien dat het zelfvertrouwen van een adolescent toeneemt terwijl dat ook haar sociale vaardigheden toenemen, en we willen weten of het een uit het ander voortkomt. Er kan ook sprake zijn van een onderliggende cyclus: Binnen de dag nemen positieve gevoelens doorgaans eerst toe en dan weer af, gedurende de week is er ook een cyclus, waarbij negatieve emoties pieken op maandag, en voor sommige mensen is er ook duidelijk sprake van een maandcyclus of een jaarcyclus.

Als we dit soort trends niet adequaat modelleren, kan dit onze resultaten behoorlijk verstoren (Liu & West, 2016; Wang & Maxwell, 2015). Maar tegelijkertijd moeten we ook weer voorzichtig zijn met het zondermeer wegfilteren van dergelijke trends en cycli: Als we de wisselwerking tussen de ontwikkeling van zelfvertrouwen en sociale vaardigheden bij adolescenten willen begrijpen, en we halen daar eerst de trend uit weg, dan hebben we nu juist datgene weggehaald waar we het meest in geïnteresseerd zijn. Maar hoe het dan wel moet, is nog steeds een open vraag.

Een andere vraag is hoe we processen die zich op verschillende tijdschalen af spelen met elkaar in verband kunnen brengen (Hamaker & Wichers, 2017). Opvoeding is iets wat zich afspeelt over een periode van 18 jaar, en tegelijkertijd is het iets wat zich afspeelt op een schaal van seconden: Hoe reageer ik als een van mijn kinderen zijn handen niet wil wassen voor het eten, of boos is omdat hij niet langer mag playstationen? Hoe kunnen we de cumulatie van dit soort micromomenten betekenisvol relateren aan de volwassene die na 18 jaar wordt afgeleverd?

Een derde vraag is hoe we de onderliggende causale verbanden kunnen onderzoeken met longitudinale data. In de sociale wetenschappen leren we over het algemeen dat we niet over causaliteit mogen praten als we geen experiment hebben gedaan, maar dat maakt ons er niet minder nieuwsgierig naar. In plaats van om de hete brij heen te draaien en allerlei vermijdende terminologie zoals associatie, invloed, en effect te gebruiken, is de wetenschap meer gebaat bij het expliciet benoemen van onze interesse in causale mechanismen, en het in kaart brengen van de voorwaarden om tot wetenschappelijk kennis daaromtrent te komen (Hernán, 2018). Hierbij moeten we ook zeker kijken hoe er in andere vakgebieden, zoals de epidemiologie en de econometrie, met dit soort vraagstukken wordt omgegaan, en welke tools daar zijn ontwikkeld om op grond van niet-experimenteel onderzoek toch meer grip te krijgen op de onderliggende causale structuur.

## **We staan op een kantelpunt.**

Het arsenaal aan statistische methodes is enorm toegenomen en groeit nog steeds. Ook de dataverzameling zal steeds meer mogelijkheden bieden. Hoe moet een onderzoeker uit al deze mogelijkheden kiezen?

Wat we hiervoor nodig hebben is een nieuwe methodologisch kader voor longitudinaal onderzoek. Wat een centrale rol zou moeten krijgen binnen dit kader is hoe we theorie, dataverzameling en data-analyse op elkaar kunnen afstemmen (Collins, 2006). Dit lijkt misschien nogal een open deur, maar in de praktijk blijken onderzoekers vaak pas echt over de analyses en hun onderzoeksvraag na te gaan denken als de data al verzameld zijn.

Mijn ideaal is dat onderzoekers beginnen met de theorie, en daaruit heel concreet proberen af te leiden welke aspecten van een proces ze nu precies willen onderzoeken. Enerzijds gaat het er dan om hoe het proces het beste gemeten kan worden: moet dat met zelfrapportage of met fysiologische metingen, en moet dat met meerdere metingen per seconde of is het juist een proces dat zich afspeelt op een schaal van minuten of uren. Anderzijds is het van belang te weten welke analysetechniek aansluit bij deze metingen en de onderzoeksvraag. Hierbij is het essentieel om aan te geven welk modelaspect nu eigenlijk interessant is: gaat het om de variantie, om een trend of cyclus, gaat het om een relatie tussen twee variabelen over de tijd, of om de samenhang binnen hetzelfde tijdstip?

Wanneer we deze aanpak proberen te gebruiken, zullen we waarschijnlijk vaak tot de ontdekking komen dat bestaande theorieën te weinig details bevatten om dit soort concrete zaken direct uit af te kunnen leiden. Dit is niet vreemd, want tot voor kort was het domweg niet mogelijk om processen zo nauwkeurig te meten, en was er dus geen enkele noodzaak om expliciet te zijn over dergelijke aspecten. Ik hoop dat de nieuwe mogelijkheden op gebied van dataverzameling en data-analyse daarom ook zullen leiden tot nieuwe theorieontwikkeling waarin dit soort aspecten

van processen, zoals de tijdschaal waarop zij zich afspelen en de specifieke patronen die zij genereren, een prominenter rol krijgen.

Zo een methodologisch kader is niet van de een op de andere dag ontwikkeld, en het zal de inzet en samenwerking vergen van enerzijds methodologen, statistici en psychometrici, en anderzijds inhoudelijke onderzoekers. We zullen hiervoor ook een nieuwe taal moeten ontwikkelen, een terminologie die het mogelijk maakt om met elkaar te communiceren over modellen en processen.

Het is te verwachten dat Nederland en Vlaanderen een belangrijke bijdrage zullen leveren aan deze ontwikkelingen: Dankzij het werk van Adriaan de Groot (De Groot, 1961), bestaat er een zeer sterke traditie van methodologie in de lage landen. Daar wordt in andere landen soms met enige jaloezie naar gekeken: methodologie is daar aan de meeste universiteiten niet als een volwaardige discipline ontwikkeld. Het wordt er over het algemeen door inhoudelijke onderzoekers een beetje bij gedaan. Maar om echt vernieuwend bezig te kunnen zijn op het gebied van methoden en statistiek, is het noodzakelijk om je daar ook echt in te specialiseren.

Het vereist ook de bereidheid van universiteiten, overheden, en de samenleving om in dit soort vernieuwing te investeren zonder de keiharde garantie dat het ook wat gaat opleveren. Toen ik 15 jaar geleden promoveerde, zag vrijwel niemand dat de vraag naar de specialistische technieken waarin ik me had bekwaamd zó enorm zou toenemen; ook ik zag dat niet. We kunnen niet weten wat de toekomst brengt, en als we wachten met investeren totdat we dat wel weten, zijn we al te laat. Als wij koplopers willen voortbrengen, moeten we jonge onderzoekers de ruimte geven om een dergelijke voorsprong te ontwikkelen.

Ik hoop dat ik u vandaag heb kunnen overtuigen van het feit dat de toegepaste statistiek niet alleen een razend interessant en levendig vakgebied is, maar ook dat innovaties op dit gebied essentieel zijn als wij een van onze kerndoelen binnen de sociale wetenschappen willen verwezenlijken: het ontcijferen van het individu. Hieraan zal ik ook met mijn leerstoel Longitudinal Data Analysis een bijdrage blijven leveren.

## Tot slot.

Nu ik het einde van mijn oratie nader, wil ik graag kort stil staan bij het feit dat de leerstoel die ik nu bekleed een *Westerdijkleerstoel* is, vernoemd naar Johanna Westerdijk. Zij sprak in 1917 in deze zaal haar oratie uit en werd daarmee de eerste vrouwelijke hoogleraar van Nederland. De Westerdijkleerstoelen worden ook wel roze leerstoelen genoemd, omdat het een initiatief is om het aantal vrouwelijke hoogleraren in Nederland toe te laten nemen. Daar wordt soms wat besmuikt over gedaan, en dat vind ik jammer. Ik ben juist erg trots—en dankbaar—in een land te leven waarin wij ons inzetten om mannen en vrouwen gelijke kansen te bieden. Dat is nog steeds hard nodig, en geen vanzelfsprekendheid, ook niet in onze westerse wereld.

Dit brengt mij bij mijn dankwoord. Allereerst wil ik het College van Bestuur, de decaan, en de beoordelingsadviescommissie bedanken voor het in mij gestelde vertrouwen.

Ook wil ik mijn promotor Peter Molenaar en mijn co-promotor Conor Dolan hartelijk bedanken. Met mijn achtergrond in klinische psychologie was ik buitengewoon ongeschikt voor het project waarop ik werd aangenomen. Ik ben hen beiden zeer dankbaar voor de bijzondere kans die ze mij hebben geboden en alles wat zij mij hebben geleerd; het heeft mijn leven wezenlijk veranderd, en het voor mij mogelijk gemaakt om werkzaam te zijn op een vakgebied waar ik echt gepassioneerd over ben.

Ik wil ook alle collega's van het *Departement Methoden en Statistiek* van onze universiteit bedanken voor het scheppen van een prettige werkomgeving. In het bijzonder dank ik professor Herbert Hoijtink en professor Peter van der Heijden voor de ruimte die ze mij in al die jaren hebben gegeven om mijn eigen onderzoekslijn te ontwikkelen.

Een speciaal woord van dank vandaag wil ik wijden aan Raoul Grasman. Hij heeft mij over de jaren heen ontzettend veel geleerd over tijdreeksanalyse, R, en latex; zonder zijn hulp en geduld was er weinig van mij terecht gekomen, vrees ik.

Daarnaast wil ik Arnold van Emmerik bedanken; jouw talent om vooral niet te zwaar te tillen aan het leven is voor mij vreemd en verfrissend. Jarenlang hebben wij in het diepste geheim aan een paper over de brieven van Vincent van Gogh gewerkt; we fantaseerden wel eens dat we daarover geïnterviewd zouden worden bij de Tros Nieuwsshow of De Wereld Draait Door. Inmiddels is het artikel meer dan anderhalf jaar geleden verschenen; ik heb vanochtend nog even gekeken en het is nog steeds nul keer geciteerd. Maar zoals de Boeddha al zei: het gaat om de reis, niet om het einddoel.

Ook wil ik op deze plek heel graag Eveline Crone bedanken, niet alleen voor het prachtige Westerdijkleerstoelen-initiatief van *Athena's Angels*, maar vooral ook voor haar warme en trouwe vriendschap sinds het begin van onze aio tijd: wat heerlijk om alle successen en teleurstellingen van het academische bestaan en het leven daarbuiten met jou te kunnen delen.

Lieve ouders, of het nu nature of nurture is, zonder jullie had ik hier niet gestaan. Wat fijn dat jullie er vandaag bij zijn.

Tenslotte mijn eigen dynamische systeem: Naut, Brent, en Dave. Steeds als ik denk dat ik het snap, is alles ineens weer anders; de wiskundige in mij worstelt daarmee; de psycholoog is vooral gefascineerd; maar bovenal heb ik jullie lief, zo lief.

Ik heb gezegd.

## Referenties

- Asparouhov, T., Hamaker, E. L., & Muthén, B. (2018). Dynamic structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*, 359-388.
- Borkenau, P., & Ostendorf, F. (1998). The Big Five as states: How useful is the Five Factor model to describe intraindividual variations over time? *Journal of Research in Personality*, *32*, 202–221.
- Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, *16*, 5-17.
- Borsboom, D., & Cramer, A. O. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, *9*, 91–121.
- Carver, C. S., & White, T. L. (1994). "Behavioral Inhibition, Behavioral Activation, and Affective Responses to Impending Reward and Punishment: The BIS/BAS Scales". *Journal of Personality and Social Psychology*, *67*, 319–333.
- Cattell, R. B. (1952). The three basic factor-analytic research designs—their interrelations and derivatives. *Psychological Bulletin*, *49*, 499-520.
- Cattell, R. B., Cattell, A. K. S., & Rhymer, R. M. (1947). P-technique demonstrated in determining psychophysiological source traits in a normal individual. *Psychometrika*, *12*, 267-288.
- Collins, L. M. (2006). Analysis of longitudinal data: The integration of theoretical model, temporal design, and statistical model. *Annual Review of Psychology*, *57*, 505-528.
- Cowles, M. (1989). *Statistics in psychology: An historical perspective*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- De Groot, A. (1961). *Methodologie: Grondslagen van onderzoek en denken in de gedragswetenschappen*. Den Haag: Mouton.
- Desrosières, A. (1998). *The politics of large numbers*. Cambridge: Harvard University Press.
- Geschwind, N., Peeters, F., Jacobs, N., Delespaul, P., Derom, C., Thiery, E., . . . Wichers, M. (2010). Meeting risk with resilience: High daily life reward experience preserves mental health. *Acta Psychiatrica Scandinavica*, *122*, 129–138.
- Hamaker, E. L. (2012). Why researchers should think “within-person”: A paradigmatic rationale. In: M. R. Mehl & T. S. Conner (Eds.). *Handbook of Research Methods for Studying Daily Life*, 43-61, New York, NY: Guilford Publications.
- Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*, 102-116.
- Hamaker, E. L., & Wichers, M. (2017). No time like the present: Discovering the hidden dynamics in intensive longitudinal data. *Current Directions in Psychological Science*, *26*, 10–15.
- Hamaker, H. C. (1961). Toegepaste Statistiek. Inaugurele Rede 1960. *Statistica Neerlandica*, *15*, 1-18.
- Hernán, M. A. (2018). The C-word: Scientific euphemisms do not improve causal inference from observational data. *American Journal of Public Health*, *108*, 616-619.

- Houben, M., Van den Noortgate, W., & Kuppens, P. (2015). The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychological Bulletin*, *141*, 901–930.
- Keijsers, L. (2016). Parental monitoring and adolescent problem behaviors: How much do we really know? *International Journal of Behavioural Development*, *40*, 271–281.
- Kuppens, P., Allen, N. B., & Sheeber, L. B. (2010). Emotional inertia and psychological maladjustment. *Psychological Science*, *21*, 984-991.
- Kuppens, P., Sheeber, L. B., Yap, M. B. H., Whittle, S., Simmons, J. G., & Allen, N. B. (2012). Emotional inertia prospectively predicts the onset of depressive disorder in adolescence. *Emotion*, *12*, 283–289.
- Liu, Y., & West, S.G. (2016). Weekly cycles in daily report data: An overlooked issue. *Journal of Personality*, *84*, 560-579.
- Luborsky L. (1995). The first trial of the P technique in psychotherapy research – A still lively legacy. *Journal of Consulting and Clinical Psychology*, *63*, 6–14.
- Molenaar, P. C. M. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika*, *50*, 181–202.
- Ormel, J., & Schaufeli, W. B. (1991). Stability and change in psychological distress and their relationship with self-esteem and locus of control: A dynamic equilibrium model. *Journal of Personality and Social Psychology*, *60*, 288-299.
- Rogosa, D. R. (1980). A critique of cross-lagged correlation. *Psychological Bulletin*, *88*, 245–258.
- Wang, P., & Maxwell, S. E. (2015). On disaggregating between-person and within-person effects with longitudinal data using multilevel models. *Psychological Methods*, *20*, 63–83.
- Weldon, W. F. R., Pearson, K, & Davenport, C. B. (1901). Editorial. *Biometrika*, *1*, 1-2.
- Wichers, M., Barge-Schaapveld, D. Q., Nicolson, N. A., Peeters, F., de Vries, M., Mengelers, R., & van Os, J. (2009). Reduced stress-sensitivity or increased reward experience: The psychological mechanism of response to antidepressant medication. *Neuropsychopharmacology*, *34*, 923–931.
- Wichers, M., Peeters, F., Geschwind, N., Jacobs, N., Simons, C. J., Derom, C., . . . van Os, J. (2010). Unveiling patterns of affective responses in daily life may improve outcome prediction in depression: A momentary assessment study. *Journal of Affective Disorders*, *124*, 191–195.