

University of Groningen

In the absence of a gold standard

Noordhof, Arjen

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2010

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Noordhof, A. (2010). *In the absence of a gold standard*. s.n.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Nederlandse samenvatting

Als er geen gouden standaard bestaat

In de kern gaat het proefschrift over de (on)mogelijkheid om objectief vast te stellen of en in welke mate iemand aan een psychische stoornis lijdt: het ontbreken van een gouden standaard. Grofweg behandel ik hierbij twee interessante en gerelateerde vraagstukken uit de psychodiagnostiek. Ten eerste de vraag naar 'de latente structuur van psychopathologie' en ten tweede de vraag naar de meetbaarheid van psychopathologie. In de eerste hoofdstukken gaat het met name om de latente structuur. Latente structuur staat hier tegenover observaties. Uit observaties die gedaan worden willen we uitspraken doen over 'wat het achterliggende probleem is'. Bij bepaalde symptomen hoort (volgens de Diagnostic Statistical Manual, DSM-IV) de diagnose 'Major Depressive Disorder' (MDD) en bij andere symptomen hoort de diagnose 'Generalized Anxiety Disorder' (GAD). De suggestie die dit geeft is dat er twee stoornissen bestaan, MDD en GAD, en dat sommige mensen die stoornis hebben en anderen niet. Dit 'er bestaan twee stoornissen...' is een uitspraak over de latente structuur van psychopathologie. In dit geval een uitspraak die veelvuldig betwist wordt en die ook niet volgt uit een nauwkeurig toepassen van DSM-IV. Het is maar zeer de vraag of iemand bij wie de diagnose MDD en de diagnose GAD zijn vastgesteld, lijdt aan twee psychische stoornissen. Ook voor veel andere diagnoses in DSM-IV zijn zeer weinig aanwijzingen dat ze zouden gaan om 'discrete stoornissen': ziekten die je wel of niet hebt. Een veelvuldig aangeprezen alternatief is om mensen in te schalen op een aantal 'dimensies' ('hoog depressief', 'behoorlijk teruggetrokken', 'matig agressief', etc.). Dit lost op zich het probleem van latente structuur niet direct op, want ook de uitspraak 'er bestaan drie dimensies van psychische problemen' kent vele problemen. Het zorgt echter wel voor een relativering van de pretentie 'ziekten' te diagnosticeren. Daarnaast zijn er methoden om deze dimensies zodanig te kiezen dat symptomen die in de algemene bevolking vaak tegelijkertijd voorkomen, zoals 'piekeren' en 'slapeloosheid', onder dezelfde dimensie vallen ¹. Deze dimensies zijn niet direct observeerbaar en worden daarom latente variabelen genoemd. Een model met verschillende dimensies heet daarom een 'latente variabelen model'. Het samen voorkomen van symptomen wil natuurlijk niet zeggen dat ze uitingen zijn van 'hetzelfde', maar het is plausibel dat wanneer symptomen vaak samengaan ze in een bepaald oorzakelijk verband staan en dus 'met elkaar te maken hebben'. Het vaak samen gaan van symptomen wordt uitgedrukt in hun covariantie en latente modellen worden zodanig gekozen dat ze zoveel mogelijk van deze covariantie 'verklaren'.

¹ Voor de volledigheid: het is niet zo dat dergelijke modellen alleen met dimensies kunnen worden gemaakt, maar in de praktijk worden meestal dimensies gebruikt en werkt dat goed.

In het tweede hoofdstuk heb ik² een heel aantal van dergelijke latente variabelen modellen gemaakt en getoetst in hoeverre ze passen bij de verzamelde data. In het onderzoek TRAILS (Tracking Adolescents' Individual Lives Survey) wordt een grote groep jongeren (2230) gevolgd sinds het einde van de basisschool. Inmiddels zijn de meesten van de middelbare school af en is de vierde meting (T4) al een heel eind gevorderd. Ikzelf heb meegewerkt aan de derde meting en gebruik in mijn onderzoek vragenlijstgegevens van de eerste drie metingen (T1-T3). Het onderzoek dat ik in het tweede hoofdstuk presenteer is gebaseerd op door ouders gerapporteerde klachten. Specifiek heb ik een bestaand en veel gebruikt dimensioneel model (het model van internaliserende en externaliserende problemen) uitgebreid met klachten die daar tot nu toe niet in opgenomen waren. Dit zijn klachten die veelvuldig voorkomen bij kinderen met de diagnoses 'autistische stoornis', 'Asperger's syndroom' en 'PDD-NOS'³. Het bleek goed mogelijk om deze klachten op te nemen in het model, maar daarvoor moest het model worden uitgebreid met een dimensie die specifiek te maken heeft met 'autisme-achtige problemen', in het Engels: 'The broader autism phenotype (BAP)'. Dit resultaat zou in de praktijk kunnen worden toegepast, onder meer door bij kinderen die nu 'PDD-NOS'-ers heten een meer genuanceerde beschrijving van problemen op meerdere dimensies te geven. Overigens betwijfel ik of de in hoofdstuk twee gebruikte term 'autism phenotype' wel zo geschikt zou zijn voor de praktijk, aangezien het woord autisme, bijvoorbeeld in films, zeer specifieke en zelfs stigmatiserende associaties oproept die bij veel van deze kinderen helemaal niet aan de orde zijn.

Het derde hoofdstuk gaat verder in op de praktische toepassing van deze dimensionele modellen in de klinische praktijk. Met name gaat het in dit hoofdstuk om de vraag of het mogelijk en nuttig is om toch categorieën uit dimensionele scores te maken. Ik laat zien hoe dit zou kunnen en bespreek verschillende doelstellingen waarvoor dit nuttig zou zijn. De belangrijkste zijn communicatie (het communiceert niet makkelijk met getallen op 4 dimensies) en het nemen van beslissingen (moet iemand nu wel of niet een diagnose en behandeling krijgen). Wat betreft het communicatieprobleem zijn er bruikbare manieren om van dimensionele scores hanteerbare taal te maken. Het beslissingsprobleem ligt veel ingewikkelder en wordt in hoofdstuk 3 slechts gedeeltelijk opgelost. Enerzijds wordt besproken dat bij veel beslissingen lokale informatie van groot belang is en het dus niet per se optimaal is om hier internationaal vaste criteria voor te schrijven. Anderzijds wordt gesteld dat het voor sommige maatschappelijke doeleinden nodig is dat experts beslissen wie wel en wie niet behandeling (vergoed) zouden moeten krijgen.

² Ik gebruik in deze samenvatting steeds de ik-vorm, maar het werk voor ieder hoofdstuk is steeds een samenwerking met meerdere auteurs geweest.

³ In Nederland wordt meestal deze Engelse afkorting van Pervasive Developmental Disorder – Not Otherwise Specified gebruikt, dus pervasieve ontwikkelingsstoornis – Niet Anderszins Omschreven.

Voor deze expertbeslissingen zou het verstandig kunnen zijn om algemene criteria vast te leggen en te evalueren in hoeverre goed geïnformeerde experts met elkaar tot overeenstemming kunnen komen. Dit blijft echter een heikel punt, des te meer door de grote verschillen tussen verschillende bronnen van informatie.

De hoofdstukken 4 tot 6 gaan over deze grote verschillen. Het blijkt dat verschillende informanten (leraar, ouder, zelf) die dezelfde vragenlijsten over hetzelfde kind invullen tot heel verschillende rapportages komen. Deze conclusie komt voort uit het observeren van covariantie (zie boven) tussen rapportages van verschillende informanten op dezelfde vragenlijstitems. Het blijkt dat heel verschillende probleemgebieden (bv. agressief en teruggetrokken gedrag) gerapporteerd door eenzelfde informant vaak duidelijk méér covariantie vertonen dan hetzelfde probleemgebied gerapporteerd door verschillende informanten. Het maakt dus voor de schatting van de mate van psychische problemen nogal wat uit aan wie je om informatie vraagt.

In hoofdstuk 4 heb ik een poging ondernomen om de informatie van verschillende informanten (zelf, leraar, ouder) over verschillende probleemgebieden (agressie, teruggetrokken gedrag, angst en meer) te integreren in één dimensioneel model. De cruciale assumptie waarop dit model stoelt is dat de meeste verschillen tussen informanten worden veroorzaakt door systematische verschillen tussen het perspectief van waaruit informanten naar een persoon kijken en de context waarin zij met die persoon te maken krijgen. Informanten zou je moeten selecteren op basis van de mate waarin ze overeenkomen én verschillen wat betreft perspectief en context. Leraar en ouder komen bijvoorbeeld deels overeen wat betreft dit perspectief, maar verschillen juist wat betreft context (school versus thuis). Vandaar dat in dit model de systematische verschillen tussen zelf-rapportage en rapportage door leraar en ouder wordt aangeduid met de term *Perspectief*, terwijl systematische verschillen tussen leraar en ouder worden aangeduid met de term *Context*. De zo gevormde componenten *Context* en *Perspectief* blijken samen een zeer belangrijk deel van de variantie in de gerapporteerde scores te 'verklaren'. Hieruit blijkt opnieuw de grote invloed van de gekozen informanten op de uiteindelijke meting. Er wordt bovendien de belangrijke stap gezet om na te denken over waarom informanten van elkaar verschillen. Een heel andere manier om over deze verschillen na te denken komt aan de orde in hoofdstuk 5.

In hoofdstuk 5 stel ik enigszins provocerend de vraag aan de orde of covariantie tussen zeer verschillende probleemgebieden (externaliserend: agressie, delinquentie, etc. en internaliserend: teruggetrokken gedrag, angstig, depressief, etc.) wel echt bestaat (fact) of dat het slechts voortkomt uit de methode van meten (artefact). Het is een nogal technisch artikel, waarin ik een vijftal methodische problemen bespreek en laat zien dat met name de gebruikte informanten en instrumenten invloed hebben op het inschatten van de associatie tussen deze verschillende probleemgebieden. Uiteindelijk kon ik de hypothese dat het slechts om een artefact gaat niet verwwerpen. Erg waarschijnlijk vind ik het overigens niet dat het 'enkel een artefact' is. In de

discussie bij dit hoofdstuk bespreek ik dat het onderscheid tussen fact en artefact in het geval van structureel verschillende informanten niet goed gemaakt kan worden. Dat wat iemand zelf rapporteert kun je niet simpelweg als 'artefact' beschouwen, maar is eerder het resultaat van een specifieke interactie tussen iemands eigen ervaringen en herinneringen, de aangeboden vragen en de manier waarop iemand besluit tot een bepaalde respons (0, 1 of 2).

Toch blijft het merkwaardig dat verschillende metingen van 'hetzelfde' tot zulke verschillende resultaten leiden. In hoofdstuk 6 stel ik daarom dat we moeten kiezen: ofwel de metingen meten dezelfde dimensie, maar dan meten ze die wel bijzonder onnauwkeurig. Ofwel, de metingen meten verschillende dimensies, maar dan is het eigenlijk vreemd om deze verschillende dimensies dezelfde naam (bijvoorbeeld agressie) te geven. Wat ik verder laat zien in hoofdstuk zes is dat de verschillen tussen zelf-rapportage en rapportage door een ouder behoorlijk stabiel zijn. Dat wil zeggen dat als de ouder op de basisschool meer, of juist minder, rapporteert dan het kind, dit verschil waarschijnlijk gedurende de adolescentie zal blijven. Wel is het zo dat zelf-rapportage gemiddeld enigszins verandert in de richting van ouder-rapportage: als een ouder meer problemen rapporteert dan is er een grotere kans dat zelf-rapportage zal toenemen. Dit effect is echter klein en het is maar de vraag of dit gegeven erg bruikbaar is in de klinische praktijk. De vuistregel is dat je kunt verwachten dat verschillen blijven bestaan.

Dit betekent dat we bij twee informanten te maken hebben met twee behoorlijk stabiele, maar toch zeer verschillende schattingen van psychische problemen. In het laatste hoofdstuk heb ik een poging gewaagd om uit deze resultaten te komen tot een manier om met informatie van meerdere informanten om te gaan wanneer er geen gouden standaard bestaat. Twee keuzes zijn hierbij cruciaal. Ten eerste kies ik ervoor om zelf-rapportage als een referentiemethode te beschouwen. Een referentiemethode is in zekere zin het tegenovergestelde van een gouden standaard. Waar een gouden standaard het onbetwistbare eindpunt van twijfel en discussie markeert, is een referentiemethode juist het betwistbare beginpunt daarvan. Betwistbaar, omdat ik er niet vanuit ga dat iemand zelf een uitermate goede beoordelaar is van zijn of haar eigen psychologische conditie. Beginpunt, omdat de zelf gevoelde symptomen het punt van vertrek en de motivatie (zouden moeten) zijn voor het starten van hulpverlening. De bewuste eigen ervaring is dus het (onvervangbare) beginpunt waartegen nieuwe perspectieven van andere informanten en van behandelaars kunnen worden afgezet. Stap 1 in mijn voorstel betreft daarom de interpretatie van zelf-rapportage als zelf-rapportage, en dus niet als directe meting van pathologie. De tweede keuze is dat ik een expliciet onderscheid maak tussen een kwantitatieve en een kwalitatieve interpretatie van verschillen tussen zelf-rapportage en rapportage door andere informanten. Dat wil zeggen: ofwel andere informanten meten hetzelfde, maar de metingen verschillen *kwantitatief*, dat wil zeggen ze zijn niet betrouwbaar. Ofwel andere informanten meten *kwalitatief* iets anders. In stap 2 van mijn voorstel worden verschillen kwantitatief geïnterpreteerd. Dit komt er met name op neer dat de

onnauwkeurigheid van meting expliciet wordt in de rapportage. Een manier waarop dit zou kunnen is door in plaats van diagnoses (bv. ziekte wel/niet aanwezig) of exacte scores (bv. neuroticisme = 9) gebruik te maken van een interval aangeduid met onnauwkeurige termen uit de natuurlijke taal (bv. 'mild tot behoorlijk depressief', 'laag tot gemiddeld angstig'). In stap 3 worden verschillen kwalitatief geïnterpreteerd. Ik geloof echter niet dat er op dit moment duidelijke interpretatieregels hiervoor beschikbaar zijn. Dat wil zeggen dat deze interpretatie in de klinische praktijk neer zou komen op een gesprek waarin wordt gepoogd te begrijpen hoe verschillen in het individuele geval tot stand zijn gekomen.