

Inleiding

HET KLASSIEKE ZORGDENKEN VERSUS EVIDENCE-BASED DENKEN

Als we geconfronteerd worden met een zorgprobleem of een zorgvraag bij een patiënt,¹ hebben we de neiging om terug te grijpen naar een aantal ‘oude’ technieken of hulpmiddelen. Het zijn klassieke manieren voor probleemoplossing die al vele jaren gebruikt worden en die als het ware overgedragen worden van generatie op generatie. Deze lijken doelmatig te zijn, maar als we nauwkeuriger kijken naar de resultaten die ze opleveren, kunnen we toch wel belangrijke vraagtekens plaatsen. Is dit wel de beste methode om van probleem tot oplossing te komen...?

Een eerste methode die we gebruiken, is **routine**. We hebben de neiging om een handeling die we al lange tijd uitvoeren, te beschouwen als vanzelfsprekend. Worden we geconfronteerd met een situatie die we herkennen, dan grijpen we terug naar deze ‘routineuze’ aanpak. Nochtans is dit geen goede aanpak... Routine geeft het risico dat je ter plaatse blijft trappelen en geen gebruik kunt maken van nieuwe inzichten en technieken. Het is in feite stilstaan in een tijd waarin op wetenschappelijk vlak grote sprongen gemaakt worden... en de facto is het dus continu achteruitgaan.

routine

¹ Niet bij alle zorgberoepen wordt de term ‘patiënt’ gebruikt. Sommigen spreken van de cliënt, de zorgvrager, de bewoner. Om de leesbaarheid van dit boek te verhogen, wordt echter altijd de term ‘patiënt’ gebruikt.

kennis

Een tweede hulpmiddel dat we hanteren, is **kennis**. We zijn (of worden) allen zorgprofessionals die zowel tijdens de opleiding als tijdens bijscholingsmomenten vrijwel continu kennis opdoen. En de kans is reëel dat we deze kennis vroeg of laat ook kunnen gebruiken om een antwoord te bieden op een concrete zorgvraag. Het is trouwens onze morele (en zelfs wettelijke) plicht om kwalitatief hoogstaande zorg te verlenen.

ervaring

Een volgend middel waar we op kunnen terugvallen als we een beslissing moeten nemen met betrekking tot zorg, is **ervaring**. Gedurende de jaren die je als zorgprofessional op de teller hebt, ben je in contact gekomen met een veelheid aan situaties, problemen en oplossingen. En mogelijk kunnen deze ervaringen helpen om een oplossing te bieden voor ons zorgprobleem.

intuïtie

Vervolgens speelt ook altijd onze **intuïtie** mee als we een beslissing nemen. Naast ons logisch denken is er namelijk zoiets als een 'buikgevoel'. Je kunt er niet echt de vinger op leggen, maar het uit zich als een positief of negatief aanvoelen van een situatie. En dit zal, zonder dat we dit echt bewust gebruiken, onze beslissing kleuren of meebepalen.

volgzaamheid

En ten laatste speelt ook een vorm van **volgzaamheid**. We hebben in meer of mindere mate de neiging om formele leiders (vb. de hoofverantwoordelijke, de arts) en informele leiders (vb. de oudere collega) te volgen in wat deze ons opdragen. Dit kan ertoe leiden dat we er niet echt bij stilstaan, en dat is nu net waar het schoentje wringt. Dit is echter geen pleidooi voor algemene rebellie en ongehoorzaamheid. Maar bij elke situatie waar we mee geconfronteerd worden in de zorg, hebben we de plicht klinisch of therapeutisch te redeneren om tot een oplossing te komen. Het blind volgen van een opdracht staat hier haaks op.

effectiviteit

We moeten ons als zorgverstrekker (en als beleidsvoerder) afvragen of de actuele besluitvorming (met de drijfveren zoals hierboven beschreven) in de zorg wel ethisch verantwoord is. Is het wel (maatschappelijk) aanvaardbaar om zorg te verstrekken waarvan we niet zeker weten of deze werkt (**effectiviteit**)

veiligheid

efficiëntie

en of deze geen schade berokkent (**veiligheid**)? Is het wel aanvaardbaar om dure zorg te verlenen als er een goedkoper evenwaardig alternatief is (**efficiëntie**)? Met andere woorden, is deze zorg wel up-to-date en kwalitatief hoogstaand? Voor Constance Milton (zorgethicus) is het antwoord duidelijk: ‘*Als een zorgprofessional een actie uitvoert, moet er bewijs zijn dat de actie de gewenste resultaten voor de patiënt oplevert en deze een middel biedt om de kostenbeheersing in de gezondheidszorg te waarborgen.*’ [1]

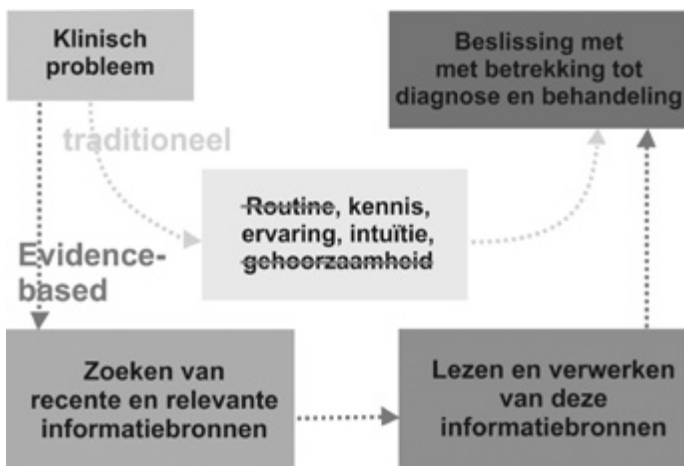
Als zorgprofessional proberen we dagelijks goede zorg te verlenen. We investeren veel tijd en middelen in deze zorg uit het oogpunt van het welzijn van onze patiënt te maximaliseren. Maar toch gebruiken we in vele gevallen nog te vaak methoden en technieken waarvoor de **wetenschappelijke onderbouwing ontbreekt**. Veel benaderingen worden immers al jaren toegepast zonder dat ze in vraag gesteld worden. De besluitvorming over het gebruik van deze methoden is vaak gebaseerd op (1) keuzes die overgedragen zijn van generatie op generatie (*tradition-based*), (2) keuzes die gemaakt zijn op basis van eensgezindheid binnen de beroepsgroep, vaak zonder wetenschappelijke rationale (*consensus-based*) en (3) keuzes die door een sociale of hiërarchische macht gemaakt zijn (*obedience-based*). De consensus- en obedience-based manier van beslissen wordt ook wel de GOBSAT-methode genoemd: *Good Old Boys Sat Around the Table*. Objectieve, kwalitatief hoogstaande informatie is ondergeschikt aan de meningen van hen die het besluit nemen. Geen enkele van deze methoden is aanvaardbaar in het moderne zorgdenken en klinisch/therapeutisch beslissen.

Bovendien wordt zorg meer en meer een multidisciplinair gegeven. Meerdere zorgverstrekkers werken rond de patiënt en bieden vanuit hun expertise een deel van de benodigde zorg aan. Het is belangrijk dat deze zorgverstrekkers op een evenwichtige en evenwaardige manier met elkaar in discussie kunnen gaan en hun activiteiten op elkaar kunnen afstemmen.

Zaken als effectiviteit, veiligheid en efficiëntie van zorg spelen hier natuurlijk sterk in mee.

We hebben dus een bijkomend hulpmiddel om nodig onze beslissing met betrekking tot zorg te onderbouwen, namelijk de kennis over effectiviteit, efficiëntie en veiligheid en die wordt ons aangereikt door de wetenschap. Dit noemt men **evidence-based onderbouwen**. Als we het echt goed willen doen, moeten we dus bij elke zorgvraag waar we tegenaan lopen gaan zoeken in recente, relevante en wetenschappelijk degelijke informatiebronnen. De oogst aan artikelen die we hierbij vinden, moeten we grondig lezen, er de essentie uit halen en deze kennis toepassen bij onze patiënt samen met een belangrijk deel van de klassieke probleemoplossing in de zorg. Routine en blinde gehoorzaamheid horen echter niet thuis binnen dit systeem.

evidence-based
onderbouwen



kritische noot

Toch moeten we hier een **kritische noot** plaatsen. De EBP-methode is, in de vorm zoals hierboven beschreven, in veel gevallen gewoonweg niet haalbaar. Het zou de zorgverstrekker enorm veel tijd kosten om informatie op te zoeken en te beoordelen... Tijd die er vaak niet is. Het zou de tevredenheid van de patiënt trouwens niet ten goede komen en het kan potentieel

gevaarlijk zijn door laattijdige zorgverstrekking. Bovendien vraagt het beoordelen van wetenschappelijke literatuur ook specifieke skills en kennis. In het verloop van dit boek zal de lezer een geheel van hulpmiddelen en structuren aangeboden krijgen waardoor het onderliggende idee van het EBP-denken en -werken wel degelijk perfect mogelijk is.

Een tweede bedenking is dat, ondanks de snelheid waarmee de wetenschap evolueert, er voor vele aspecten van de gezondheidszorg nog steeds onvoldoende (en soms zelfs volledig geen) bewijs is. Vaak moeten we dus roeien met de riemen die we hebben en ons tevredenstellen met eerder zwak wetenschappelijk bewijs.

De ladder van de beslissingen... een beetje geschiedenis

Het nemen van beslissingen is zo oud als de mensheid zelf. Vanaf het ogenblik dat we in staat waren onze instincten enigszins te overstijgen, hadden we de mogelijkheid om bewust te kiezen. De drijfveren om van probleem tot oplossing te komen zijn door de tijd heen geëvolueerd, maar alle vormen van besluitvorming zijn nog steeds aanwezig in onze maatschappij. Het is echter niet logisch of ethisch verantwoord om al deze methoden te allen tijde te hanteren in de (para)medische besluitvorming. Toch gebeurt dit spijtig genoeg nog (veel te) vaak.

TRIAL-AND-ERROR

Misschien wel de primitiefste vorm van besluitvorming en leren is de **trial-and-error methode**. Vermoedelijk hebben in de verre voorgeschiedenis van de mensheid een aantal van onze voorvaders geprobeerd om het vuur vast te grijpen. Logisch, aangezien het vuur beschermend en de warmte comfortabel was. Een deel van deze durvers heeft zich vermoedelijk flink verbrand, terwijl anderen leerden dat ze de brandende stok beter aan de 'koude' kant konden vastnemen. Idem dito voor het gebruik van vuurstenen die, als ze tegen elkaar geslagen werden, vonken genereerden waarmee ze vuur konden maken. Trial-and-error is belangrijk geweest in de ontwikkeling van de mens, maar er was nog een stap nodig, namelijk de kennisoverdracht.

EXPERIENCE-BASED

Het was belangrijk dat kennis ook langdurig opgeslagen, hergebruikt en verder gegeven kon worden. Deze *experience-based* laag is een bijkomende laag in de besluitvorming, waardoor de mens minder en minder het risico liep om met falen (error) geconfronteerd te worden. Het herinneren en delen van al dan niet positieve effecten van een handeling, was belangrijk voor zowel het individu als de groep. Om terug te grijpen naar het voorbeeld hierboven: eenmaal een persoon of een groep ervaren had hoe vuur verplaatst kon worden en deze kennis overgedragen kon worden, gaf dit een enorme evolutionaire voorsprong ten opzichte van andere groepen.

experience-based

TRADITION-BASED

Kennis over positieve en negatieve ervaringen werd gecommuniceerd en gedeeld tussen stamgenoten en overgedragen van ouder op kind. Hierdoor ontstond een reeks van **gewoontes en gebruiken** binnen de groep. Deze *tradition-based* overdracht is een belangrijk cultureel element, waardoor elk stamlid bepaalde 'regels' in acht kon nemen en de overgedragen kennis kon gebruiken, ter meerwaarde van de gehele groep. Het creëerde een vorm van zekerheid en stabiliteit voor de groep, maar had ook een nadeel. Wanneer een traditie eenmaal geworteld was in de groep, was er vaak weinig intentie en animo om hieraan te tornen. Pogingen om een traditie te herzien, konden immers als bedreigend worden beschouwd voor de stabiliteit van de stam en konden leiden tot spanningen en conflict. Dit impliceert dat tradities vaak onterecht behouden bleven, hoewel nieuwe kennis aantoonde dat de overgedragen informatie achterhaald was.

tradition-based

obedience-based

OBEDIENCE-BASED

Een ander element dat speelt in het nemen van beslissingen is **macht**. Personen die al dan niet hoger geplaatst zijn op de hiërarchische ladder (zowel formele als informele leiders), kunnen een beslissing doordrukken of anderen verplichten om een bepaald gebruik aan te nemen. Van de anderen van de groep wordt verwacht dat ze de beslissing volgen. Deze *obedience-based* methode heeft voordelen in bepaalde gevallen (vb. militaire operaties), maar kan ook zeer nadelig zijn (despotisme), aangezien de beslissing mogelijk als enige doel heeft de macht te bestendigen. Als de 'leider' bovendien niet openstaat voor innovatie (die misschien niet in zijn of haar voordeel is), kan deze methode resulteren in stilstand (wat zich in de snel evoluerende medische wereld vertaalt in achteruitgang).

science-based

SCIENCE-BASED

In de achttiende eeuw ontstond een nieuwe intellectuele stroming in Europa: de verlichting. Deze was een reactie op het dogmatische (obedience- en tradition-based) geloof als basis van kennis. De verlichting vertrok van het idee dat kennis verkregen kan worden via de rede en door te filosoferen. Of anders gesteld: door zo goed mogelijk gebruik te maken van rationele, redelijke en kritische vermogens om problemen en vraagstukken met betrekking tot het menselijk bestaan aan te pakken. [2] Hier werd de basis gelegd voor datgene wat wij verstaan onder 'wetenschap', namelijk een menselijke activiteit om de wereld op een zo objectief mogelijke wijze te onderzoeken en betrouwbare kennis te vergaren. Manieren om dit te bekomen, zijn het uitvoeren van observaties en metingen, het opzetten van experimenten, het analyseren van de verzamelde gegevens, het opstellen van hypothesen en het controleren van elkaars bevindingen. Zo wordt geprobeerd een robuuste *body of knowledge* (corpus van kennis) over allerhande

kennisgebieden op te bouwen. De overtuiging is dat diegene die zich baseert op deze objectieve kennis voor het nemen van beslissingen (*science-based*), de objectiefste methode gebruikt om tot een oplossing te komen.

science-based

Kwaliteit van wetenschappelijk onderzoek

Wereldwijd worden een immens aantal onderzoeken gevoerd, naar een veelheid van onderwerpen, in een brede waaier van onderzoeksgebieden. Wetenschappelijk onderzoek is geen sinecure. Het vraagt een gedegen expertise en is gebonden aan een groot aantal methodologische spelregels. De belangrijkste kwaliteitscriteria van wetenschappelijk onderzoek zijn: onafhankelijkheid, herhaalbaarheid, toetsbaarheid, generaliseerbaarheid en bruikbaarheid. [3, 4] Deze criteria worden hieronder verder toegelicht. Een overkoepelende eigenschap van onderzoek is transparantie. Elke stap in het proces moet immers verifieerbaar zijn door externe partijen.

KWALITEITSCRITERIA VAN WETENSCHAPPELIJK ONDERZOEK ^[5]

onafhankelijk-
heid

Onafhankelijkheid: de methode, analyse, interpretatie en conclusie van een onderzoek mogen niet beïnvloed worden door niet-wetenschappelijke drijfveren. Er mag geen invloed zijn van betrokkenen (stakeholders), financiers, voorkeuren of strekkingen, of van de subjectieve mening van de onderzoeker zelf. Onderzoek dient intersubjectief te zijn. Dit betekent dat onderzoekers het met elkaar eens moeten zijn over de degelijkheid van de opzet van de studie en over de gevonden resultaten.

herhaalbaarheid

Herhaalbaarheid: door de nauwkeurige beschrijving van elke stap van de methodologie moet het mogelijk zijn voor

anderen om het onderzoek te herhalen (replicatie). Dit zou in principe moeten leiden tot vergelijkbare resultaten. Replicatie van onderzoek draagt bij aan de overeenstemming van onderzoekers over de resultaten en aan de betrouwbaarheid.

Toetsbaarheid: de uitspraken die een wetenschapper doet over zijn onderzoek en de conclusies met betrekking tot zijn populatie moeten toetsbaar zijn. Vaak vertrekt een onderzoeker van een vooraf gedefinieerde hypothese die al dan niet bevestigd wordt door het lopende onderzoek. Een gedetailleerde beschrijving van het proces, vertrekkende van de opzet van het onderzoek, via de uitvoering tot de conclusie(s), is hierin zeer belangrijk.

Generaliseerbaarheid: onderzoek probeert een uitspraak te doen over een effect op of een fenomeen bij een welomschreven populatie. Deze populatie kan zelden in haar geheel betrokken worden in het onderzoek en een duidelijke aflijning is dan ook aan de orde (vb. de groep van diabetes type I-patiënten tussen twaalf en achttien jaar). Daarom is het belangrijk voor degelijk onderzoek om goed stil te staan bij de samenstelling en de grootte van de steekproef (onderzoekssample). De bevindingen bij een goed samengestelde steekproef kunnen immers binnen bepaalde grenzen veralgemeend (gegeneraliseerd) worden naar de volledige populatie. Een inaccurate samenstelling van de steekproef kan leiden tot foute conclusies en daarom tot problematische generalisering van de resultaten. Daarom is het belangrijk om zowel de steekproef (grootte, samenstelling, inclusie, exclusie) als het onderwerp en de opzet van de studie (tijd, ruimte, definitie, perspectief) zo nauwkeurig mogelijk te definiëren.

Bruikbaarheid: een wetenschappelijk onderzoek kan worden beschouwd als een stukje informatie dat past in een grote puzzel. Daarom is het belangrijk dat de resultaten van een onderzoek bruikbaar zijn voor andere wetenschappers om verder op te bouwen (fundamenteel onderzoek) of voor eindgebruikers om toe te passen in de praktijk (toegepast onderzoek).

toetsbaarheid

generaliseer-
baarheid

bruikbaarheid

BETROUWBAARHEID EN VALIDITEIT ^[5]

Kwaliteit van onderzoek kan beoordeeld worden op basis van **betrouwbaarheid en validiteit**. Elk aspect van de opzet – de uitvoering, de samenstelling van de steekproef, *response rate* (percentage deelnemers dat antwoordde op de oproep om in te stappen in het onderzoek), de datacollectie, de analyse, de interpretatie van de resultaten en de conclusies – speelt hierin een rol. Beide begrippen zijn echter niet zwart-wit, maar worden uitgedrukt als een ‘mate van’.

betrouw-
baarheid

Betrouwbaarheid ^[3, 6]

In theorie (en in de praktijk) kunnen onderzoeken toevallige fouten bevatten. Bij een controle op betrouwbaarheid wordt nagegaan in welke mate toevallige fouten voorkomen in het onderzoek. Hoe betrouwbaarder het onderzoek, hoe groter de kans dat een herhaling van het onderzoek een vergelijkbaar resultaat zal opleveren. Het optreden van fouten in een onderzoek is onvermijdelijk. Zo kan een respondent een vraag fout interpreteren en in een vragenlijst een verkeerd antwoord aanduiden. Ook bij het ingeven van data in statistische software kunnen fouten gemaakt worden. Er zijn echter methoden om dergelijke fouten te voorkomen of op zijn minst te minimaliseren. Hierna een bloemlezing van hulpmiddelen.

- **Sample size:** als de steekproef van een onderzoek groter is, is het effect van een fout antwoord bij één respondent kleiner. De ideale steekproefgrootte kan statistisch berekend worden (*sample size calculation, power calculation*).
- **Pre-test:** wanneer een vragenlijst vooraf afgenomen wordt bij een testgroep en daarna geoptimaliseerd, is de kans op interpretatie- of logicafouten kleiner.
- **Input-errors:** als meerdere onderzoekers de invoering van de data dubbelchecken en er zo weinig mogelijk overgetypt wordt (vb. met een online vragenlijst), voorkom je invoerfouten.

- **Triangulatie:** wanneer meerdere methoden van dataverzameling gecombineerd worden en de resultaten op elkaar aansluiten, is de betrouwbaarheid van het onderzoek groter.
- **Hertesten:** als een onderzoek meerdere malen herhaald wordt (replicatie) en de resultaten vergelijkbaar zijn, kun je ervan uitgaan dat de aanpak betrouwbaar is.
- **Statistische significantie:** bij kwantitatief onderzoek kan door middel van statistiek met grote zekerheid aangetoond worden of een fenomeen al dan niet aan toeval te wijten is.
- **Interne consistentie:** wanneer meerdere vragen in een vragenlijst focussen op dezelfde eigenschap (vb. mate van angst), kan op een statistische wijze aangetoond worden in hoeverre de verschillende items van het meetinstrument dat ook daadwerkelijk doen.
- **Datasaturatie:** als bij kwalitatief (niet-kwantitatief) onderzoek de dataverzameling doorgaat totdat zich geen nieuwe informatie meer aandient, kan met vrij grote zekerheid besloten worden dat de dataverzameling compleet is.
- **Transparantie:** wanneer onderzoeksmethoden en -resultaten kristalhelder weergegeven worden, kunnen lezers van het onderzoeksverslag minutieus nagaan of alles correct verlopen is.
- **Peerreview:** wanneer andere (externe) onderzoekers het onderzoeksverslag vóór de publicatie nalezen, opmerkingen formuleren om de verstaanbaarheid en correctheid te verhogen, en zich uiteindelijk akkoord verklaren, verhoogt dit de betrouwbaarheid van het onderzoek.

validiteit

Validiteit ^[6]

Wanneer je een onderzoek voert, is het belangrijk dat je daadwerkelijk het fenomeen meet dat je wilt meten. Daarom is het belangrijk dat dit fenomeen grondig beschreven is, met inbegrip van al zijn subdimensies en onderdelen. Bovendien is het belangrijk dat het instrument dat je gebruikt om je fenomeen te meten, ook echt geschikt is om dat te doen. Het gaat hier dus concreet om de mate waarin we meten wat we hopen te meten (en wat we denken te meten). Als we bijvoorbeeld de afstand tussen twee punten willen meten, is het belangrijk een goede (geijkte) meetlat te gebruiken. Maar ook als we de mate van zelfredzaamheid willen meten, hebben we een goed instrument nodig.

Voor een onderzoek maakt men onderscheid tussen interne en externe validiteit. [7]

interne validiteit

- **Interne validiteit:** een onderzoek heeft een goede interne validiteit als de verschillen tussen de onderzoeksgroepen uitsluitend aan het effect van de onderzochte interventie te wijten zijn en niet aan het meetinstrument of de studieopzet. Interne validiteit is belangrijk wanneer we uitspraken willen doen over oorzaak-gevolgrelaties, zoals in experimenteel onderzoek. De kernvraag voor interne validiteit is of het gevonden resultaat kan worden toegeschreven aan de interventie en niet aan andere mogelijke oorzaken voor dit resultaat.

externe validiteit

- **Externe validiteit:** de externe validiteit of generaliseerbaarheid refereert aan de mate waarin het resultaat van het onderzoek veralgemeenbaar is naar andere populaties dan louter de studiepoulatie. Het is de mate waarin je op basis van een steekproef uitspraken kunt doen over de volledige onderzoekspoulatie. Hier wordt soms ook gesproken over 'extrapolatie'.

De validiteit van een meetinstrument kun je vaststellen op vier gebieden. Het is echter niet altijd mogelijk om alle vier de aspecten te beschrijven.

- **Gezichtsvaliditeit** (*face validity*): dit is de eenvoudigste vorm van validiteit. Het zegt in welke mate een meetinstrument op het eerste gezicht het concept wat het zou moeten meten, ook werkelijk lijkt te meten. Niet zelden beperkt de vaststelling van de validiteit van een onderzoeksinstrument zich tot deze gezichtsvaliditeit.
- **Inhoudsvaliditeit** (*content validity*): inhoudsvaliditeit is de mate waarin een meetinstrument (vb. vragenlijsten of schalen) alle aspecten van een concept ook werkelijk meet. Enerzijds is het belangrijk dat het concept dat men wil meten zo grondig mogelijk beschreven is, met inbegrip van al zijn subdimensies. Anderzijds is het belangrijk dat bij de meting van het concept alle deelaspecten ook werkelijk bevraagd of gemeten worden. Als je een instrument ontwikkelt om de mate van depressie te meten, is het belangrijk dat ook alle kerndimensies van depressie hierin opgenomen zijn.
- **Criteriumvaliditeit** (*criterion validity*): criteriumvaliditeit is een aspect van validiteit dat nagaat in hoeverre scores op een meetschaal of instrument correleren met externe, niet-testcriteria. [8] Het uiteindelijke doel van criteriumvaliditeit is aan te tonen dat testcores voorspellend zijn voor uitkomsten in het echte leven. Zo kan een bedrijf bij aanwerving van een potentiële werknemer een test doen om zijn capaciteiten en competenties te meten. De vraag is echter in welke mate de uitkomst van deze test de werkelijke prestaties van de werknemer op de werkvloer (nu en in de toekomst) kan voorspellen.
- **Constructvaliditeit** (*construct validity*): constructvaliditeit geeft aan hoe goed een instrument meet wat het geacht wordt te meten. Met andere woorden, is het instrument zo geconstrueerd dat het succesvol test wat het beweert te testen? Constructvaliditeit wordt meestal geverifieerd

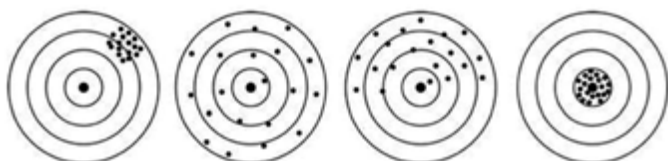
gezichtsvaliditeit

inhoudsvaliditeit

criterium-
validiteitconstruct-
validiteit

door het instrument te vergelijken met andere instrumenten die vergelijkbare kwaliteiten of eigenschappen meten om te zien hoe sterk de twee maten zijn gecorreleerd.

Zowel betrouwbaarheid als validiteit is essentieel in een onderzoek. Een goede manier om dit visueel voor te stellen, is de metafoer van de roos van een schietschijf. [9] Als alle kogelinslagen dicht bij elkaar liggen, maar verder van het centrum van de roos verwijderd zijn, is je onderzoek betrouwbaar maar niet valide. Als de inslagen verspreid zijn over de hele roos, is je onderzoek valide maar niet betrouwbaar. Enkel als de kogelinslagen dicht bij elkaar en het centrum van de roos liggen, kun je stellen dat je onderzoek valide en betrouwbaar is.



Betrouwbaar
maar niet valide

Valide maar
niet betrouwbaar

Niet valide en
niet betrouwbaar

Valide en
betrouwbaar

Verskil tussen betrouwbaarheid en validiteit

SAMENGEVAT

Wetenschappelijk onderzoek is complex en methodologisch zwaar. Het moet voldoen aan een hele reeks kwaliteitscriteria: onafhankelijkheid, herhaalbaarheid, toetsbaarheid, generaliseerbaarheid en bruikbaarheid. Betrouwbaarheid en validiteit van het onderzoek zijn bovendien essentieel.

Bedenking over de kwaliteit, betrouwbaarheid en validiteit van onderzoek

STATISTISCHE SIGNIFICANTIE EN PRAKTISCHE RELEVANTIE

Het is belangrijk om te beseffen dat niet alle resultaten van wetenschappelijk onderzoek ook een praktische meerwaarde hebben. Een onderzoeker kan via statistiek aantonen dat een interventie een effect heeft dat niet aan het toeval te wijten is (en dat dus met andere woorden statistisch significant is; p -waarde $< .05$). Maar deze significantie zegt niets over de grootte van het gevonden effect (*effect size*). Zo kan onderzoek uitwijzen dat een bepaalde vorm van organisatie van zorg een statistische toename van tevredenheid geeft bij de zorgvrager, maar als deze toename maar 2% is en de 'reorganisatie' groot, kun je de vraag stellen of het wel relevant is om deze reorganisatie door te voeren. Het spreekt voor zich dat de *effect size* veel belangrijker is voor de zorgprofessional dan de statistische significantie... Een andere methode om de praktische relevantie van een resultaat te beoordelen, is aan de hand van het **Number Needed to Treat** (NNT) of *Number Needed to Harm* (NNH). Dit is een vorm van effect- of risicoberekening, vertrekkend van observationele studies, die het effect vergelijkt van een interventie met placebo (geen actieve therapie) of een andere interventie. Hoe dan ook geeft het NNT aan hoeveel

effect size

Number Needed to Treat

Number Needed to Harm

personen je moet behandelen om één succesrijke behandeling meer te hebben dan bij placebo (of de vergelijkende therapie). Het relateert met andere woorden de gezondheidswinst van een behandeling aan het aantal patiënten dat je moet behandelen. Het NNH is de tegenhanger van het NNT en vertelt iets over de kans op bijwerkingen bij een behandeling.

STATISTISCHE SIGNIFICANTIE EN GEZOND VERSTAND

Een onderzoek kan aantonen dat er een statistisch verband is tussen twee variabelen. Maar het is ook belangrijk om na te gaan of dit verband wel logisch is (of met andere woorden of er een goede rationale is) om dit verband te verklaren. Om dit bevattelijk te maken, verduidelijk ik het graag aan de hand van een voorbeeld

In 2012 publiceerde Messerli het artikel ‘Chocolate Consumption, Cognitive Function, and Nobel Laureates’ in *The New England Journal of Medicine*. [10] Hij zette de gemiddelde chocoladeconsumptie per inwoner van 23 landen uit tegen het aantal Nobelprijzen die in deze landen gewonnen werden. Hij toonde een statistisch zeer sterk verband aan tussen chocoladeconsumptie per capita en het winnen van een Nobelprijs (correlatie: $r = 0.791$, $p < .0001$) (zie figuur op p. 31). Zijn conclusie was dat chocolade de cognitieve functie verbetert en daarom de kans op het winnen van belangrijke wetenschappelijke prijzen kan verhogen. Een bizarre maar statistisch correcte conclusie...

Het artikel was echter bedoeld om aan onderzoekers duidelijk te maken dat niet elk statistisch verband ook een logisch verband is. Chocolade was immers niet de rechtstreekse ‘oorzaak’ van het effect. Chocolade is simpelweg een luxeproduct dat hoofdzakelijk in rijke landen gegeten wordt. Rijke landen hebben meer mogelijkheden om wetenschappelijk onderzoek te financieren en dus ook een grotere kans om Nobelprijzen in