

Het is niet al goud wat er blinkt

Jaap van der Stel

Verslaving, 3, nummer 4 (2007) p. 14-17

Dat effectiviteit het sleutelwoord is bij de beoordeling van behandelprogramma's, kan niemand zijn ontgaan. Effectiviteit wordt vastgesteld in experimenten, die in de gezondheidszorg RCT's worden genoemd: gerandomiseerde en gecontroleerde proeven. Idealiter weten zowel de behandelaar als de proefpersoon niet welke behandeling ze krijgen. Bij pillen is dat goed te doen, want aan de buitenkant kun je niet zien wat er in zit. Bij psychotherapie is het natuurlijk uitgesloten dat de behandelaar niet weet wat hij doet. Hoe het ook zij, het RCT wordt alom gezien als 'gouden standaard'. Is dat terecht? En zo nee, wat verdient dan wel die status?

Eerst een paar positieve dingen. De geneeskunde en de klinische psychologie zouden nooit zo goed mensen kunnen helpen als ze het experiment links hadden laten liggen. Eeuwenlang hebben artsen 'maar wat aan gerotzooïd' - status, het aanzien van je leermeester of de statuur van het gebouw waarin de praktijk werd uitgeoefend gaven de doorslag. Ook de consultatiebureaus voor alcoholisme vestigden zich na de oorlog liefst in herenhuizen met een opgaand stoepje. Wie binnentrad was gediagnosticeerd. Moedige artsen doorbraken vanaf eind achttiende eeuw het onbewezen gezag van hun collega's. Met experimenten (inclusief controlegroepen) toonden ze aan dat citrusvruchten scheurbuik tegengaan, gewassen handen aan het kraambed verplicht zijn en dat aderlaten niet helpt. Registratie en statistiek bewezen in dat verband hun nut. Zonder experimenten en daaraan voorafgaande hypothesen bleek gestage wetenschappelijke vooruitgang onmogelijk. Ook in de geestelijke gezondheidszorg en verslavingszorg werd het experiment een leidend beginsel.

Met een experiment kunnen we vaststellen dat iets werkt. Maar op de vraag waaróm dat zo is, geeft een experiment geen antwoord. Een RCT toont aan dat een interventie bijdraagt aan een (al dan niet gewenste) uitkomst. Sceptici kunnen tevreden gesteld worden met het verhogen van de interne validiteit, door te proberen alle (bekende) variabelen 'onder controle' te krijgen. Maar helaas gaat een hoge interne validiteit ten koste van de externe validiteit: hoe zekerder je er over bent dat deze interventie, bij deze mensen, heeft geleid tot deze uitkomst, des te minder stellig kun je zijn dat zo'n uitkomst ook bij andere mensen wordt bereikt. Dat komt doordat mensen zoveel van elkaar verschillen en telkens met andere omstandigheden te maken hebben. Men probeert natuurlijk de experimenten zo vaak mogelijk te herhalen, bij uiteenlopende groepen mensen, maar het blijft behelpen.

Het is niet toevallig dat er geregeld medicijnen die alle testen hebben doorstaan van de markt worden gehaald, of dat het indicatiegebied wordt ingeperkt. Kort geleden was dat nog het geval met rimonabant, dat behalve als 'afslankpil' hoge ogen scoort bij onderzoekers in de verslavingszorg. Inmiddels is door de instanties besloten dat het medicijn bestemd is voor mensen met obesitas plus hart- en vaataandoeningen en diabetes, en dat het niet mag worden voorgeschreven aan depressieve patiënten en/of patiënten die antidepressiva gebruiken; lastig in de verslavingszorg. In Amerika heeft de FDA, de Amerikaanse autoriteit op het gebied van de toelating van geneesmiddelen, zelfs besloten dat het middel zo gevaarlijk is dat het niet mag worden verkocht.

Onderzoekers moeten veel meer aandacht besteden aan de zoektocht naar de complexe mechanismen en determinanten die aan ziekten ten grondslag liggen. En ze moeten veel

meer kennis vergaren over hoe een therapie werkt of waarom deze vaak misloopt. Dat bespaart een hoop ellende. Pas als we de mechanismen en de determinanten daarvan kennen, kunnen we ziekteprocessen verklaren en betere voorspellingen doen of, en bij wie, een bestaande behandeling wel of niet zal werken. En pas dan kunnen we werkzame interventies ontwerpen die aanspraak maken op een hoge externe validiteit.

Een RCT is een 'mechanismeloze' onderzoeksactiviteit - uiterst zinvol maar goed beschouwd slechts leidend tot oppervlakkige kennis. Dat het geldt als gouden standaard is niet terecht; die statusoekenning is volgens mij misleidend. Anders gezegd: experimenten zijn het zilver, kennis over mechanismen levert het goud. Dat de gemiddelde lezer van de talloze al dan niet goed uitgevoerde RCT's het spoor vaak bijster raakt is begrijpelijk, want vaak is de hoeveelheid data omgekeerd evenredig aan het aantal inzichten dat ermee verbonden is.

Een alternatief voor het RCT is er niet, maar enkele inherente problemen moeten wel duidelijk voor het voetlicht worden gebracht. De eis om alle variabelen onder controle te houden kan in het medisch en psychologisch onderzoek zelden ingewilligd worden. De veronderstelling dat de kenmerken van de experimentele groep (grotendeels) identiek zijn aan die van de toekomstige 'doelgroep' van de interventie is niet realistisch. En slechts bij uitzondering geldt dat een interventie in alle gevallen leidt tot de gewenste uitkomst. In werkelijkheid is slechts sprake van kansen en effectgrootten, die gunstiger of ongunstiger zijn dan datgene waarmee een interventie wordt vergeleken. Claims dat een interventie een noodzakelijke en voldoende voorwaarde is voor herstel worden zelden waargemaakt. En hoe beter een RCT wordt uitgevoerd, des te meer het zegt over de gebeurtenissen in het experiment en des te minder over iets daarbuiten. Het is dus niet terecht dat het RCT het unieke bestanddeel is van evidentie op het hoogste niveau.

Hoe vermijd je inferieure 'informatie' en bereik je superieure 'kennis'? Onderzoekers zouden (wereldwijd) het accent moeten leggen op de zoektocht naar determinanten en mechanismen van de systemen die ze willen beïnvloeden. Dat vergt de samenvatting van alle beschikbare kennis en het verwerven van nieuwe kennis, en het vormen van (netwerken van) hypothesen over de werking van de systemen en deze testen. Hypothesen waarvoor empirische steun is - en die elkaar versterken - kunnen samengebracht worden in een theorie. Op basis daarvan kan een model worden ontwikkeld over de wijze waarop het te beïnvloeden systeem zich wetmatig gedraagt - ook wanneer eigenschappen daarvan worden gewijzigd. Voorspellingen daaromtrent moeten bij herhaling worden getest. Het resultaat is kennis over de werking (en ontwikkeling) van het te beïnvloeden systeem, in alle denkbare omstandigheden. Vervolgens kunnen hypothesen over de te testen interventie worden gevormd en bijeengebracht worden in een 'handelingsmodel', dat uiteraard ook moet worden getest. De methodologische kennis over het RCT komt dan goed van pas.

Het voordeel van een werkwijze die focust op kennis over de determinanten en mechanismen van systemen - zoals een 'verslaafd brein' - is dat er minder onzekerheid bestaat of geregistreerde (gunstige of ongunstige) effecten aan de interventie mogen worden toegeschreven. Ook het probleem dat een goede controlemogelijkheid soms ontbreekt, is minder urgent. Een goed voorbeeld is het rekenmodel van het Centraal Planbureau. Het model is gebaseerd op *mechanismische* kennis van de Nederlandse economie, dat wil zeggen dat er goede hypothesen zijn over hoe interventies daarin specifieke effecten bewerkstelligen. Het voorspelt dus goed het systeemgedrag wanneer variabelen worden veranderd, zoals rijksuitgaven, het rentepercentage of de lonen. Door studie van de uitkomsten wordt het model steeds accurater - zelfs zonder dat er een

controlegroep bij aan te pas komt. Het model is verder intern valide, ofschoon het door veranderende omstandigheden wel continu bijgesteld moet worden; een universeel geldig model bestaat niet.

Ik stel voor dat we, waar het verslaving betreft, alles op alles zetten om een omvattend werkingsmodel van verslaving of verslavingsgedrag tot stand te brengen. Laten we ons even inhouden om het zoveelste praktijkexperiment in gang te zetten, waarin we nauwelijks weten volgens welke mechanismen een interventie een verandering bewerkstelligt. Want onthoud: weten dát iets werkt is zilver, weten waardóór is pas goud.