

Wie niet gemiddeld is, zien we niet; een methodologisch pleidooi voor inclusieve evidentie*

F.L. Truijens, M.M. De Smet, M. Desmet, R. Meganck

- Achtergrond** In psychologisch en psychiatrisch onderzoek worden methodologische standaarden gehanteerd voor het ontwikkelen van een evidence-base voor de klinische praktijk. Elke methode vormt 'evidentie' vanuit specifieke methodologische assumpties. De gekozen methode bepaalt zo wat telt als 'evidentie' en vormt daarmee de organisatie van de klinische praktijk.
- Methode** We bespreken kwalitatieve analyses van drie patiënt-participanten in 'gouden standaard'-psychotherapieonderzoek, die opvielen in de sample omdat zij op een uitgesproken manier met vragenlijsten omgingen.
- Resultaten** Deze 'rijke casussen' maken inzichtelijk hoe het voor gegeven nemen van methodologische assumpties waardevolle klinische informatie uit het zicht kan laten verdwijnen, waardoor de evidence-base niet vanzelfsprekend passend is.
- Conclusie** Door mensen op voorhand of tijdens de studie van analyses uit te sluiten of in het gemiddelde te verliezen, verliezen we de mogelijkheid om juist die mensen een wetenschappelijk onderbouwde behandeling te bieden. Als we evidence-based willen werken, moeten we dus ook evidentie verzamelen voor wat niet evident is.

Met de focus op 'evidence-based' werken, heeft de geestelijke gezondheidszorg zich in de laatste decennia ontwikkeld tot een professionele en wetenschappelijk onderbouwde praktijk.¹ Psychologische behandeling is bewezen effectiever dan de tijd zijn werk laten doen, en meer dan 350 specifieke therapievormen en behandelprotocollen zijn wetenschappelijk effectief bevonden.² Dat al die therapieën gemiddeld goed werken, betekent echter niet dat *alle* mensen er baat bij hebben – en onduidelijk is waarom een therapie werkt, of bij wie. Dit besef leidt veldbreed tot de roep om meer aandacht voor de vraag: 'wat werkt voor wie?' Dit heeft een weelde aan onderzoek opgeleverd naar specifieke werkzame ingrediënten, gemeenschappelijke factoren, procesfactoren, en inter- en intra-individuele predictoren.³⁻⁵ Omdat we weten dat ruim tien procent van de mensen onder behandeling verslechtert of behandeling juist als last ervaart (*treatment burden*), is deze beweging naar een meer verfijnde en gelaagde benadering niet alleen wetenschappelijk, maar vooral klinisch-ethisch van groot belang.^{6,7}

Welke evidentie is nodig?

Daar waar generieke bevindingen niet blijken te passen is niet de vraag *of* we evidentie nodig hebben, maar *welke* evidentie. In het evidence-based discours plaatst men een dergelijke vraag doorgaans in het kader van methodologische principes en – al dan niet 'gouden' – standaarden.

Zo wordt in psychologisch en psychiatrisch onderzoek veel waarde gehecht aan de 'hiërarchie van evidentie', waarin men onderzoek op populatieniveau prefereert boven casuïstiek, en waarin statistische vergelijking als uitgangspunt wordt genomen, bijvoorbeeld door vergelijking van pre- en postsymptoomniveaus, vergelijking van symptoomreductie in verschillende behandelcondities, en vergelijking van effectgroottes in meta-analyses. Deze benadering berust voor een wezenlijk deel op paradigmatische overtuigingen: elke methodologie bevat een reeks methodologische aannames en principes die op zichzelf niet aan onderzoek worden onderworpen, maar wel sterk bepalend zijn voor hoe 'evidentie' wordt gedefinieerd. Omdat de resulterende evidentie vervolgens een grote rol speelt in organisatie, uitvoering en verantwoording van de klinische praktijk, is het van belang om stil te staan bij hoe methodologie bepaalt wat als evidentie begrepen wordt. Een belangrijke noot daarbij is dat elke methodologie evidentie in een specifieke vorm aanlevert. Als goudenstandaardonderzoek uitgaat van één methodologische variant wordt evidentie logischerwijs ook tot één vorm wordt beperkt. De vraag of die vorm van evidentie ook daadwerkelijk het meest wenselijk en toereikend is voor de klinische praktijk kan echter niet puur methodologisch worden beantwoord: dit vraagt om een gezamenlijke denkoefening over de filosofische en wetenschappelijke fundamenten onder onze wetenschap en praktijk.

AUTEURS

Femke Truijens, universitair docent Klinische Psychologie & Methodologie, Erasmus Universiteit Rotterdam, Department of Psychology, Education and Child Studies; gastonderzoeker, Universiteit Gent, Faculteit Psychologie en Pedagogische Wetenschappen; psychotherapeut privépraktijk.

Melissa De Smet, universitair docent Kwalitatieve Onderzoeksmethoden, Universiteit Tilburg, Tilburg School of Social and Behavioral Sciences; postdoctoraal onderzoeker, Universiteit Gent, Faculteit Psychologie en Pedagogische Wetenschappen.

Mattias Desmet, professor Klinische Psychologie, Universiteit Gent, Faculteit Psychologie en Pedagogische Wetenschappen; psychoanalyticus privépraktijk.

Reitske Meganck, professor Klinische Psychologie, Universiteit Gent, Faculteit Psychologie en Pedagogische Wetenschappen; psychoanalyticus privépraktijk.

Correspondentie

Femke Truijens (truijens@essb.eur.nl).

Geen strijdige belangen meegedeeld.

Het artikel werd voor publicatie geaccepteerd op 26-3-2024.

Citeren

Tijdschr Psychiatr. 2024;66(6):265-269

Missen we betekenisvolle variatie?

Om de knelpunten in deze denkoefening klinisch voelbaar te maken, bespreken we drie casussen van patiënten die als patiënt-participant deelnamen aan de *Ghent Psychotherapy Study*, een goudenstandaardpsychotherapiestudie waarin de effectiviteit van psychodynamische en cognitieve gedragstherapie werd vergeleken.³ In deze studie (medisch-ethisch getoetst vanuit het Universiteitsziekenhuis Gent, registratienummer B670201318127) werden naast de goudenstandaardvragenlijsten ook interviews afgenomen voor, tijdens en 5 keer na therapie (tot 2 jaar follow-up), en werden therapie sessies op audio opgenomen.³ De statistische analyses van symptoomreductie leidden tot de conclusie dat beide therapievormen gemiddeld genomen werkzaam waren – wat overigens een veelvoorkomende bevinding is in zulk goudenstandaardtherapieonderzoek.³ Deze conclusie is gebaseerd op analyses van de volledige steekproef, waar onze drie patiënt-participanten ('June', 'Amy' en 'John') deel van uitmaakten.^{8,9} Deze participanten sprongen echter in het oog, omdat ze ieder op een particulier opvallende manier met de vragenlijsten in het onderzoek omgingen. Daarmee vormen deze cases 'rijke gevallen', die we volgens de principes van 'purposive sampling' in detail onderzochten, om aanknopingspunten te vinden die ook informatief kunnen zijn op grotere schaal.¹⁰

June, die zich aanmeldde met depressie en ernstige angstklachten, ervoer de vragenlijsten als een doorlopende confrontatie met potentiële 'zothed', of tekenen dat ze psychisch volledig 'gek' was (Citaat uit therapie sessie 40).¹¹ Het invullen van de vragenlijsten werd voor haar een stressor die haar angst- en stemmingsklachten verergerde, wat tot een hogere symptoomscore na therapie leidde. Dit effect wordt 'measurement reactivity'

genoemd,¹² en staat bekend als methodologisch probleem. Tegelijkertijd maakte deze reactiviteit inzichtelijk hoe haar angst door willekeurige woorden, die ze bijvoorbeeld opving in media of zelfs gesprekken tussen anderen, aangewakkerd kon worden.⁸ Zo gaf de reactiviteit op vragenlijsten een ingang voor therapeutisch werk en werd de achteruitgang in klachten een mogelijkheid voor therapeutische groei.

Amy startte haar therapie met vergelijkbare angst om 'zot' te zijn (Postinterview), maar ervoer de vragenlijsten juist als geruststelling.⁸ Waar zij zichzelf altijd als 'gek' had gezien, vond ze het relativerend dat bijvoorbeeld haar altijd al aanwezige suïcidale gedachten 'maar' met 2 op 4 gescoord hoefden te worden op de *Beck Depression Inventory-II*, wat haar zelfbeeld nuanceerde en haar ervaren angst sterk verminderde.^{11,13} Ook bij Amy was er dus sprake van measurement reactivity, maar met een wezenlijk andere uitwerking dan bij June. Sterker nog, de symptoomreductie die Amy ervoer, schreef zij zelf expliciet toe aan het scoren van het vragenlijsten, en dat stond voor haar los van het therapeutisch effect dat ze op andere vlakken ervaren had.

De casussen van June en Amy roepen vragen op die doorgaans in goudenstandaardtherapieonderzoek niet worden opgemerkt. Onderzoekers gaan er normaliter vanuit dat bij gebruik van een degelijk gerandomiseerde en gecontroleerde (RCT-)onderzoeksopzet en gevalideerde symptoomvragenlijsten, data automatisch valide zijn om de causale vraag naar therapie-effectiviteit te beantwoorden.^{14,4} Binnen deze goudenstandaardbenadering wordt een serie expliciete en impliciete aannames gedaan die zelden worden gecheckt, zoals de aanname van generaliseerbaarheid.⁴

Onze goudenstandaardmethode gaat uit van *gemiddelden* in gedrag en gedragsverandering, waarbij wordt aan-

genomen dat als de steekproef random getrokken werd en groot genoeg was, die gemiddelden generaliseerbaar zijn van de ‘representatieve steekproef’ naar de populatie waar die steekproef uit getrokken werd. De casussen van June en Amy herinneren ons er echter levendig aan dat een gemiddelde niet wil zeggen dat de meerderheid rond het midden scoort: het kan even goed zijn dat het gemiddelde wordt getrokken uit mensen die heel negatief of heel positief scoren. De standaarddeviatie wordt normaliter gebruikt om te controleren of die variatie statistisch acceptabel is, maar *inhoudelijk* wordt er verder niet meer naar gekeken. Hoewel op basis van statistische en methodologische principes dus verdedigbaar is dat gemiddelden representatief en generaliseerbaar zijn, blijkt bij verdieping in de inhoud en ervaringen van participanten die representativiteit niet vanzelfsprekend houdbaar.

Statistische ‘uitsluiting’

Een ander, en klinisch uiterst relevant voorbeeld van hoe methodologische assumpties bepalend zijn voor de praktijk, is hoe ‘error’ wordt gedefinieerd. In de statistische methoden die we in goudenstandaardonderzoek gebruiken, wordt er vanuit gegaan dat er een zekere orde in data zit. Als bepaalde metingen op een onlogische manier van deze orde afwijken, wordt dat beschouwd als afwijking, uitschieter, of error. In de statistiek is het gebruikelijk om de dataset op te schonen door zulke afwijkingen uit de analyses te verwijderen. Maar de consequentie is dat wie afwijkt, nooit in de ‘evidence-base’ zal verschijnen. Mensen met complexe, chronische, of comorbide problematiek, of mensen die niet typisch reageren op de behandelprotocollen, maken deel uit van onze dagelijkse klinische praktijk, maar worden in onze goudenstandaardmethoden feitelijk als ‘error’ beschouwd en buiten de evidence-base geplaatst. Om dit te illustreren, presenteren we John, patiënt-participant in de *Ghent Psychotherapy Study*.³ Ook hij ervoer een therapeutisch effect dat sterk verweven bleek met reactiviteit op de vragenlijsten voor het onderzoek. Waar hij vol enthousiasme aan het onderzoek was begonnen met de verwachting dat hij aan de universiteit eindelijk echt goede therapie zou krijgen, voelde hij zich gedesillusioneerd en zelfs ‘gebruikt’ (Therapiesessie 4), omdat de onderzoeksdata die hij via de vragenlijsten aanleverde niet *in* de therapie gebruikt werden.¹¹ Dit maakte hem zo boos dat hij de dataverzameling saboteerde, door instructies opzettelijk niet of verkeerd op te volgen of vragenlijsten geheel leeg te laten. De ontbrekende data in zijn vragenlijsten waren echter niet ‘missing at random’; het weglaten van scores was immers een zeer betekenisvolle act, die bovendien inzicht gaf in symptomen waar hij al lang mee worstelde. Omdat hij de onderzoekers bovendien wilde bewijzen dat hij ‘zijn belofte wél na kwam’ (Postinterview), zette hij de therapie voort waar hij normaalgesproken – naar eigen zeggen – na de eerste twee of drie sessies zeker mee gestopt zou zijn. En tot zijn eigen verrassing had de uitgezeten therapie een

positieve impact op zijn klachten en zelfbeeld, wat werd bevestigd in de symptoomreductie die te zien was in zijn scores over tijd. Een mooi therapeutisch effect, zouden we willen concluderen, dat aansluit bij de conclusie op groepsniveau dat de therapie in kwestie effectief is. Probleem is echter dat een aanzienlijk deel van zijn data ontbrak, wat onderzoekers normaliter beweegt om een casus als deze als ‘error’ aan te merken en uit te sluiten van analyses. Maar hoe terecht is zulke exclusie in dit geval, als er wel degelijk een therapeutisch effect werd gesorteerd? En wat verliezen we, als de ontbrekende of geïnvalideerde data *juist* betekenisvolle informatie geven over de casus, zoals wat deze persoon in therapie gehouden heeft terwijl hij normaal geneigd is af te haken? Terwijl het vanuit statistische logica verdedigbaar zou zijn om John als ‘error’ weg te zetten en zijn gegevens niet mee te nemen in het groepsgemiddelde, maakt zijn casus duidelijk hoe statistische uitsluiting ook leidt tot verlies van klinisch zeer relevante informatie.

Methodologie bepaalt wat evidentie is

Zo onderstrepen deze drie casussen dat methodologische keuzes en voorkeuren niet alleen wetenschappers aangaan, maar ook in belangrijke mate de klinische professional. Methodologie *vormt* wat als evidentie wordt beschouwd, en bepaalt daarmee de kaders voor evidence-based gezondheidszorg. Zo komt methodologie binnen in de klinische praktijk, met invloed op het handelen van de professional, maar ook op de manier waarop de zorgvrager benaderd wordt. Dit sluit aan bij de al langer klinkende kritiek op het op voorhand uitsluiten van participanten met comorbide of complexe problematiek omdat die complexiteit de causale uitspraken over therapeutische effecten bemoeilijkt. De keerzijde van deze strakke inclusiecriteria is echter dat de gevonden evidentie maar beperkt representatief is voor de echte populatie die in de klinische praktijk passeert, waardoor een substantieel deel van de klinische populatie niet vanuit een passende evidence-base behandeld kan worden.^{4,11}

Dat geldt echter evenzo voor mensen *binnen* de wetenschappelijke steekproeven, die als uitschieters alsnog uit de data verdwijnen (zoals John) of waarvan interpersoonlijke variatie over het hoofd wordt gezien (zoals bij June en Amy). Normaal gaan onderzoekers ervan uit dat individuele verschillen in een willekeurig (random) getrokken steekproef wel uitmiddelen, aangenomen dat die verschillen niet systematisch, maar willekeurig zijn. Dit is opnieuw een aanname die voortkomt uit statistische methoden, maar eigenlijk niet of nauwelijks wetenschappelijk wordt gecontroleerd.¹¹ Het gevolg van deze aanname is dat ervan wordt uitgegaan dat variatie, mits deze binnen de statistisch acceptabele grenzen valt, geen verdere inhoudelijke verdieping behoeft. Door mensen op voorhand of tijdens de studie van analyses uit te sluiten of in het gemiddelde te verliezen, verliezen we echter de mogelijkheid om juist die mensen een wetenschappelijk onderbouwde behandeling te bieden. Hier stuiten we

op de paradox dat de evidence-base die ons in staat zou moeten stellen om zonder onderscheid en subjectieve voorkeur zorg te ontwikkelen en te bieden, juist selectiviteit en zelfs uitsluiting in de hand kan werken.

De zoektocht naar betekenisvolle evidentie

Als we echt evidence-based willen werken in de geestelijke gezondheidszorg, zullen we dus ook evidentie moeten verzamelen voor wat als gegeven wordt beschouwd. Dat betekent: juist wat *niet evident* is onderzoeken.

Dat vraagt ten eerste om aannames in onze standaardmethoden daadwerkelijk te onderzoeken, omdat het niet bij evidence-based werken past om belangrijke assumpties simpelweg voor waar aan te nemen. Zeker als zo'n aanname mogelijk klinisch relevante informatie toedekt, zoals de aanname dat interpersoonlijke variatie wel zal uitmiddelen en als zodanig representatief is voor een populatie vol variatie, is het van belang om die aannames empirisch te onderzoeken. Dat gebeurt bijvoorbeeld in het Meaningful Measurement (MEANS) lab (opgezet door de eerste auteur van dit artikel, zie <https://osf.io/y2fbc/>), waar met mixed methoden wordt onderzocht hoe mensen überhaupt vragenlijsten scoren, en of het inderdaad valide is om aan te nemen dat individuele processen van begrijpen en betekenisgeven cijfers opleveren dit rechtstreeks te middelen en generaliseren zijn. Ten tweede vraagt dat om het onderzoeken van mensen die normaal niet in onderzoek terechtkomen vanwege inclusiecriteria, of de mensen die als uitschieters uit onderzoek vallen, of mensen bij wie het protocol niet aanslaat.¹⁵ Dat begint met het onderzoeken van de individuen die opvallen, om zicht te krijgen op de variaties die in de 'echte wereld' buiten het laboratorium aan de orde van de dag zijn. Met systematisch casuonderzoek kunnen we niet alleen op individueel niveau opvallende contextuele zaken in kaart brengen, maar ook over gevallen heen kijken of variatie een bepaalde systematiek vertoont of toch – zoals doorgaans eenvoudigweg wordt aangenomen – individueel en toevallig is. In het eerste geval is het vanzelfsprekend aan de wetenschapper om deze serieus te nemen in aanpalend onderzoek. In het tweede geval heb je de aanname van willekeurigheid wetenschappelijk onderbouwd, waardoor de uitvoer van statistische analyses op groepsniveau gelegitimeerd is; al ontslaat ook dat scenario ons niet van de wetenschappelijke en klinische taak om de uitzondering met dezelfde toewijding en precisie te onderzoeken als het gemiddelde.

Dit argument voor onderzoek naar wat niet evident is, betekent niet noodzakelijk dat we alleen nog onderzoek op individueel niveau moeten doen, en is dus geen pleidooi tegen aggregatie. Aggregatie verwijst naar het samennemen van data om op groepsniveau patronen te kunnen identificeren. In de psychologie zijn we sterk geënt op kwantitatieve aggregatie: we verzamelen graag en vaak (voldoende) grote datasets waarin we met *probability statistics* patronen zoeken, zoals correlaties tussen theoretisch afgebakende variabelen.

In dit statistisch-methodologische kader is het logisch om 'therapie-effect' te conceptualiseren in termen van statistische significantie, en om menselijke variatie te operationaliseren vanuit een normaalverdeling.¹⁶ Het is echter even goed mogelijk om narratieve data, zoals interview- of focusgroepdata, getranscribeerde therapie sessies, ervaringsverhalen of uitgeschreven gevalsstudies, te combineren tot een dataset en met systematische kwalitatieve methoden te analyseren.^{10,17,18} Hierbij werken we ook met aggregatie van data, maar zonder vooronderstellingen over gemiddelden op populatieniveau, normaalverdelingen, statistische significantie, enzovoorts.

Hetzelfde geldt voor statistische analyses op individueel niveau, waarin aggregatie niet automatisch betrekking heeft op het aantal mensen in de steekproef, maar op het aantal meetmomenten per persoon. Zo staat analyse van herhaalde metingen, bijvoorbeeld verzameld met dagelijkse 'experience sampling'-methoden, toe om *binnen* personen patronen bloot te leggen.¹⁹ In beide voorbeelden (respectievelijk kwalitatieve en individuele statistische methoden) wordt principieel uitgegaan van het individu als vertrekpunt, om vanaf daar 'bottom-up' op zoek te gaan naar patronen die klinisch informatief en relevant kunnen zijn.

De hermeneutische benadering

Deze oriëntatie op klinisch onderzoek dat zowel methodologisch solide als klinisch relevant is, sluit aan bij ons voorstel om klinisch onderzoek als inherent *hermeneutisch* te begrijpen.^{9,11} In de hermeneutiek worden mensen beschouwd als 'meaning making agents': we zijn doorlopend bezig met betekenis, door onszelf en de wereld om ons heen te interpreteren en daar naar te handelen, wat vervolgens betekenis weer in beweging brengt.²⁰ In de klinische praktijk zijn we inherent gericht op betekenis: met welk betekenis kader komen mensen binnen in onze spreekkamer? Hoe werken of functioneren die betekenissen voor de persoon? Hoe kunnen we stagnerende betekenissen ombuigen door ze aan te vullen of uit te dagen via een therapeutisch discours? Hoe bewaken we vervolgens dat mensen dat discours als eigen discours blijven of gaan beschouwen, zonder dat het een geïnstitutionaliseerd zelf-verhaal wordt? Omdat de klinische praktijk inherent narratief is, sluit de hermeneutische benadering daar naadloos bij aan.²¹

Voorbij de cijfers

De hermeneutische benadering geeft ook een andere kijk op de cijfers waar we in onderzoek, en in toenevende mate ook in de praktijk, zo op gebrand zijn. Deze cijfers worden in ons veld doorgaans verzameld via zelf-rapportagematen, waarin we participanten vragen om hun ervaringen vanuit hun betekenis kader te vertalen in cijfers, die wij als onderzoekers of therapeuten vervolgens vanuit ons eigen betekenis kader interpreteren (d.w.z. *double hermeneutics*). Het uitdrukken in maat en getal ('becijferen') van ervaringen door participanten,

en het interpreteren, valideren en toepassen van die cijfers door professionals, kunnen we begrijpen als een ‘hermeneutische cirkel’.^{11,22} In die cirkelbeweging tussen betekenissen van patiënten en professionals, zijn cijfers net als woorden *betekenisvolle* eenheden binnen het klinische verhaal.⁸

De hermeneutische benadering is inherent bottom-up, wat wil zeggen dat ze uitgaat van de betekenis die een individu aan zichzelf en de omgeving geeft, van waaruit gedeelde betekenisgeving tussen en over mensen begrepen kan worden. Deze betekenisvolle benadering maakt het ook mogelijk om het concept ‘therapie werkt’ conceptueel te (her)overwegen.¹⁷ Waar therapeutische werkzaamheid vanuit de standaard statistische definitie als symptoomreductie wordt begrepen, kan de hermeneutische benadering in kaart brengen hoe mensen zelf ‘therapie-effect’ ervaren (d.w.z. *lived experience*), en genuanceerd inzicht geven in waar het werkt, waar het *begint* te werken, maar nog niet op stoom is (wat cruciaal is voor behandelplanning en -bijsturing), en waar het niet (meer) werkt.^{6,22} Daarin wordt ‘(nog) niet werken’ van therapie niet als *error* gezien, maar als normaal onderdeel van een therapeutisch proces. Waar clinici doorlopend te maken krijgen met individuele betekenisgeving in veranderingsprocessen, stellen wij voor een methode te hanteren die van zulke betekenisgeving vertrekt en omhoog werkt. Langs die weg kunnen we zoveel mogelijk mensen voorzien van een evidence-based praktijk die uitgaat van nuance, in plaats van een norm. Want als we niet meer zien wat niet gemiddeld is, zorgt methodologie voor een evidence-base die patiënten die niet in de pas lopen buitenspel zet.

NOOT

*Dit artikel is gebaseerd op het artikel Truijens FL, De Smet MM, Meganck R, Desmet M. Evidence for the non-evidenced: an argument for integrated methods and conceptual discussion on what needs to be evidenced in psychotherapy research in *Philosophy, Psychiatry, & Psychology* (2021; 28: 137-40). Afgedrukt met toestemming.

LITERATUUR

- Wampold B, Imel Z. The great psychotherapy debate. The evidence for what makes psychotherapy work (2de ed.). New York: Routledge; 2015.
- Kazdin A. Treatment as usual and routine care in research and clinical practice. *Clin Psychol Rev* 2015; 42: 168-78.
- Meganck R, Desmet M, Van Nieuwenhove K, e.a. The Ghent Psychotherapy Study: A pragmatic, stratified, randomized parallel trial into the differential efficacy of psychodynamic and cognitive-behavioral interventions in dependent and self-critical depressive patients. *Psychother Psychosom* 2023; 92: 267-78.
- Westen D, Novotny C, Thompson-Brenner H. The empirical status of empirically supported psychotherapies: Assumptions, findings, and reporting in controlled clinical trials. *Psychol Bull* 2004; 130: 633-7.
- Cuijpers P, Reijnders M, Huibers M. The role of common factors in psychotherapy outcomes. *Annu Rev Clin Psychol* 2019; 15: 207-31.
- Lambert M. What have we learned about treatment failure in empirically supported treatments? Some suggestions for practice. *Cogn Behav Pract* 2011; 18: 413-20.
- Demain S, Goncalves A, Areia C, e.a. Living with, managing and minimising treatment burden in long term conditions: a systematic review of qualitative research. *PLoS One* 2015; 10: e0125457.
- Truijens FL, Van Nieuwenhove K, De Smet M, e.a. How questionnaires shape experienced symptoms. A qualitative case comparison study of questionnaire administration in psychotherapy research. *Qual Res Psychol* 2022; 19: 806-30.
- Truijens F, De Smet M, Vandevoorde M, e.a. What is it like to be an object of research? On meaningful measurement and validation of data in psychotherapy research. *Methods in Psychology* 2023; 8: 100118.
- Willig C. Introducing qualitative research in psychology. Berkshire: McGraw-Hill; 2008.
- Truijens FL, De Smet M, Desmet M, e.a. Validity of data as precondition of evidence in psychotherapy research. A methodological analysis of what is taken to count as evidence in psychotherapy research. *Philosophy, Psychiatry & Psychology* 2021a; 28: 115-28.
- French D, Sutton S. Reactivity of measurement in health psychology: How much of a problem is it? What can be done about it? *Br J Health Psychol* 2010; 15: 453-68.
- Beck A, Steer R, Brown G. Beck Depression Inventory (2de ed.). San Antonio: Harcourt Assessment; 1996.
- Chambless D, Ollendick T. Empirically supported psychological interventions: controversies and evidence. *Annu Rev Psychol* 2001; 52: 685-716.
- Krivzov J, Notaerts L, Van Nieuwenhove K, e.a. The lack of failure reports in published psychotherapy case studies: Implications for dis-‘illusioning’ of research and practice. *European Journal of Psychotherapy & Counselling* 2021; 23: 139-55.
- De Smet M, Meganck R, De Geest R, e.a. What ‘good outcome’ means to patients: understanding recovery and improvement in psychotherapy for major depression from a mixed methods perspective. *J Couns Psychol* 2019; 67: 25-39.
- Iwakabe S, Gazzola N. From single-case studies to practice-based knowledge: Aggregating and synthesizing case studies. *Psychother Res* 2009; 19: 601-11.
- Levitt H, Motulsky S, Wertz F. Recommendations for designing and reviewing qualitative research in psychology: promoting methodological integrity. *Qual Psychol* 2017; 4: 2-22.
- Myin-Germeyns I, Oorschot M, Collip D. Experience sampling research in psychopathology: Opening the black box of daily life. *Psychol Med* 2009; 39: 1533-47.
- Gadamer H. Truth and method. New York: Crossroad Publishing; [1960]1989.
- Greenhalgh T. Narrative based medicine in an evidence based world. *BMJ* 1999; 318: 323-5.
- Smith J, Eatough V. Interpretative phenomenological analysis. In Lyons E, Coyle A, red. *Analysing qualitative data in psychology*. Londen: Sage; 2007.
- Kious B. Evidence for doubting the evidence? *Philosophy, Psychiatry & Psychology* 2021; 28: 129-31.
- McLeod J. Why it is important to look closely at what happens when therapy clients complete symptom measures. *Philosophy, Psychiatry & Psychology* 2021; 28: 133-6.
- Truijens FL, De Smet M, Meganck R, e.a. Evidence for the non-evidenced: An argument for integrated methods and conceptual discussion on what needs to be evidenced in psychotherapy research. *Philosophy, Psychiatry & Psychology* 2021b; 28: 137-40.