

# Herstelschaal *Individual Recovery Outcomes Counter* (I.ROC): scoring op universele schalen

E. de Beurs, M.J. Metz, L.M.W. Nahar-van Venrooij

- Achtergrond** In dit artikel bieden wij normscores voor de I.ROC, een instrument voor het meten van herstel. Normgegevens voor dit instrument op basis van de algemene bevolking worden gepresenteerd in de vorm van twee universele schalen: T-scores en percentielrangordescores (PR-scores). We bespreken voor- en nadelen van beide schalen en gaan in op hun onderlinge relatie.
- Methode** We gebruikten gegevens van een omvangrijke steekproef uit de algemene Nederlandse bevolking om PR- en T-scores te berekenen. Er werden twee benaderingen vergeleken waarmee ruwe scores naar T-scores worden omgezet: een simpele lineaire formule en een benadering waarbij ruwe scores eerst zijn genormaliseerd (rankit) en vervolgens een formule is bepaald voor de curvilineaire relatie van ruwe scores met T-scores.
- Resultaten** De frequentieverdeling van ruwe scores op de I.ROC was nagenoeg normaal en een lineaire formule voldeed voor de meeste ruwe scores. Alleen bij zeer lage scores vonden we forse verschillen tussen lineaire en genormaliseerde T-scores. We bieden een oversteektabel, -formules en een figuur om ruwe scores om te zetten naar T- en PR-scores.
- Conclusie** De I.ROC is een bruikbaar instrument dat past bij een brede visie op herstel. Voor een begrijpelijke presentatie van de uitkomsten voor zowel de cliënt als professional, en voor het gebruik van verschillende vragenlijsten naast elkaar, bevelen we toepassing van universele schalen aan. Voor incidentele omzetting van de ruwe score naar een T-score voldoet een simpele lineaire formule; voor inbouw in scoringsoftware bevelen we de precieze en complexe formule voor genormaliseerde T-scores (curvilineair) aan.

De *Individual Recovery Outcomes Counter* (I.ROC) is in 2007 ontwikkeld door Penumbra Mental Health in Schotland als instrument voor het meten van 'herstel'. Instrumenten voor het meten van herstel of 'recovery' komen voort uit de herstelbeweging.<sup>1</sup> Een kerngedachte is dat we bij het meten van behandeluitkomsten (en bij het behandelen zelf) verder moeten kijken dan symptoomreductie.<sup>2</sup> Symptoomreductie wordt ook wel 'klinisch herstel' genoemd. Naast klinisch herstel is ook het functioneren op interpersoonlijk, sociaal en maatschappelijk vlak, en het welbevinden van belang. Persoonlijk herstel omvat aspecten van het leven die ook wel worden samengevat met het CHIME-acroniem, wat staat voor connectie, hoop & optimisme, identiteit, *meaning* (betekenis) en *empowerment*.<sup>3</sup> In Nederland vinden we dit gedachtegoed terug in de positieve psychologie en in het model van Huber voor (positieve) gezondheid.<sup>4,6</sup> Voortvloeiend uit deze zienswijze op gezondheid en herstel is de I.ROC ontwikkeld. De meetpretentie van de I.ROC is dan ook breed: met de I.ROC wordt in 12 items voor diverse levensgebieden het klinisch, sociaal-maatschappelijk en persoonlijk herstel vanuit cliëntperspectief in beeld gebracht. Het instrument wordt

ingevuld door de cliënt (zelfrapportage) of door de cliënt en behandelaar/begeleider samen.<sup>7</sup> De I.ROC werd ontwikkeld in nauwe samenwerking met professionals en cliënten. Het meetinstrument is onderdeel van een meer uitgebreide methodiek, waarbij ook concrete en haalbare doelen voor de behandeling worden geformuleerd op basis van de I.ROC-uitkomsten op de diverse levensgebieden. Er is tevens een stappenplan beschikbaar om deze behandeldoelen te realiseren.<sup>7</sup>

Over de psychometrische kenmerken van de I.ROC is internationaal gepubliceerd door diverse auteurs. Vanuit Penumbra in Schotland vond het eerste validatieonderzoek plaats.<sup>7,9</sup> Daarna volgde de Nederlandse vertaling en validatie-onderzoek naar de Nederlandse I.ROC.<sup>10-12</sup> De validatieonderzoeken in Schotland en Nederland lieten positieve resultaten zien ten aanzien van de betrouwbaarheid (interne consistentie en test-hertest) en validiteit (hoge samenhang met andere meetinstrumenten en acceptabele gevoeligheid voor verandering). Uit (nog ongepubliceerde) gegevens van GGZ Breburg (schriftelijke mededeling, dr. M.J. Metz, 2024) weten we dat de bij een directe vergelijking de gevoeligheid voor verandering van de I.ROC niet onderdoet voor die van de *Symptom*

## AUTEURS

**Edwin de Beurs**, hoogleraar Klinische psychologie, Universiteit Leiden en senior onderzoeker, Arkin GGZ, Amsterdam.

**Margot Metz**, senior onderzoeker en verpleegkundige, GGZ Breburg en Tranzo Academische Werkplaats voor Zorg en Welzijn, Tilburg University.

**Lenny Nahar-van Venrooij**, senior onderzoeker en epidemioloog, Jeroen Bosch Ziekenhuis, Den Bosch/Tranzo Academische Werkplaats Kwaliteit van Huisarts en Ziekenhuiszorg, Tilburg University.

## Correspondentie

Prof. dr. Edwin de Beurs (Edwin.de.Beurs@arkin.nl).

Geen strijdige belangen meegedeeld.

Het artikel werd voor publicatie geaccepteerd op 1-5-2024.

## Citeren

Tijdschr Psychiatr. 2024;66(6):356-361

*Questionnaire (SQ-48)*.<sup>13</sup> Dit komt overeen met onderzoek waarin men een toename in totaalscore op de I.ROC rapporteerde na 10-22 weken behandeling met een effectgrootte (ES) van 0,44, meer dan met de *Recovery Assessment Scale (RAS)*; ES = 0,38).<sup>11,14</sup> Een artikel is in voorbereiding met resultaten van vervolgonderzoek naar de betrouwbaarheid, factorstructuur en convergente validiteit van de I.ROC, gebaseerd op gegevens van de ggz-populatie. Normgegevens voor de Nederlandse I.ROC zijn echter nog niet voorhanden. Met dit artikel willen wij voorzien in deze normgegevens, uitgedrukt op universele schalen in de vorm van T-scores en percentielrangordscores (PR-scores), met de Nederlandse bevolking als referentiegroep. Een artikel over de betrouwbaarheid, factorstructuur, de convergente validiteit en de gevoeligheid voor verandering is in voorbereiding.

## Waarom T- en PR-scores?

We gebruiken talloze zelfrapportagemeeinstrumenten in de ggz en elk meetinstrument heeft zijn eigen schaal. Het resultaat wordt uitgedrukt in een ruwe score (de somscore of de gemiddelde score van de antwoorden op vragen) en dit leidt tot een unieke schaal voor elk meetinstrument. Het vereist kennis en enige ervaring met het instrument om gevoel te krijgen bij een score: is de score gemiddeld, hoog of laag? De ruwe score zegt eigenlijk pas iets wanneer we die vergelijken met scores van een referentiegroep, zoals de algemene bevolking of cliëntengroepen. Voor de juiste interpretatie van de meetresultaten publiceren vragenlijstontwikkelaars vaak normtabellen, waarin je op kunt zoeken hoe gewoon of uitzonderlijk een score is. Normtabellen delen ruwe scores op in klassen (bijvoorbeeld: zeer hoog, hoog, boven gemiddeld, gemiddeld, beneden gemiddeld, laag, zeer laag). Dit kan beter en verfijnder. Eerder hebben we gepleit voor het gebruik van twee universele schalen: T-scores en PR-scores. Beide schalen drukken met een getal uit hoe uitzonderlijk of hoe gewoon een score op een meetinstrument is.<sup>15-17</sup> Ze komen zo in de plaats van de gebruikelijke normering van scores in veel grovere

scoreklassen. De T-score is uit te leggen aan cliënten als 'deze score geeft met een standaardmaat de plek van uw resultaat weer ten opzichte van een vergelijkingsgroep van de Nederlandse bevolking'. De PR-score kan uitgelegd worden als: 'Deze score geeft aan hoeveel procent van de vergelijkingsgroep een lagere score heeft dan u'. Een belangrijk verschil tussen T- en de PR-scores is dat T-scores een intervallschaal hebben, terwijl PR-scores slechts volgordelijk (ordinaal) zijn. Dit impliceert dat je T-scores van elkaar mag aftrekken, bijvoorbeeld om een getalswaarde toe te kennen aan de mate van vooruitgang tijdens behandeling (Delta T). Bij PR-scores is dat niet toegestaan.<sup>18</sup> De PR-score is makkelijker te begrijpen, maar de T-score is breder te gebruiken. Bij intelligentietests en in het onderwijs zijn deze universele schalen al heel gebruikelijk; denk bijvoorbeeld aan de IQ-score, een gestandaardiseerde score met een gemiddelde van 100 en een standaarddeviatie van 15. En de PR-score kennen we onder meer van de Eindtoets Basisonderwijs, beter bekend als de CITO-toets.<sup>19</sup>

Brede invoering van universele schalen maakt de interpretatie van de uitslag, de vergelijking met andere instrumenten en de communicatie met collega's én cliënten over uitkomsten van meetinstrumenten eenvoudiger. Dit is relevant voor de praktijk van de ggz, zowel voor de professional als voor de cliënt. Tijdens behandeltrajecten worden, via *routine outcome monitoring (ROM)*, vragenlijsten over klachten en herstel periodiek afgenomen. Hiermee kunnen de cliënt en de professional samen de behandeling evalueren en afspraken maken over vervolgstappen.<sup>20-21</sup> In toenemende mate hebben cliënten ook zelf inzage in de ROM-uitkomsten. Een uniforme en daarmee beter te begrijpen weergave van de uitkomsten van meetinstrumenten, zoals van de I.ROC, wordt hiermee steeds belangrijker.<sup>17</sup>

In dit artikel vergelijken we, op basis van I.ROC-data die werden verzameld in een representatieve steekproef van de algemene Nederlandse bevolking, twee benaderingen om T-scores te verkrijgen. We doen een voorstel voor omzetting van ruwe I.ROC-scores naar universele schalen.

## METHODE

### Steekproef

Voor de dataverzameling maakten we gebruik van het LISS-panel (Langlopende Internet Studies voor Sociale wetenschappen) van Centerdata.<sup>22</sup> In een eerdere studie (preprint) is bij een aselechte steekproef van 3246 LISS-panelleden de I.ROC uitgezet;  $n = 2459$  hebben de I.ROC volledig ingevuld (75,8%).<sup>23</sup> 88 personen deden dat binnen 3 minuten, wat kan wijzen op erg vluchtig antwoorden (*satisficing questionnaires*).<sup>24</sup> De data van deze respondenten werden uit het bestand verwijderd, waarmee er 2371 overbleven. Gegevens over sekse of leeftijd ontbraken bij nog eens 5 personen, waardoor 2366 respondenten resteerden met complete gegevens. Van de respondenten waren 1274 vrouw (53,8%). De gemiddelde leeftijd was 53,9 jaar (SD: 17,7; uitersten: 18-94). Voor een vergelijking van leeftijdsgroepen werden 3 ongeveer even grote subgroepen onderscheiden: 18-46, 46-65 en 66-94 jaar. Uitgebreide informatie over de steekproef is te vinden via Centerdata ([www.lissdata.nl](http://www.lissdata.nl)).<sup>22</sup>

### Meetinstrument, de I.ROC

De items van de I.ROC worden per item op een scherm of als een pagina in een boekje gepresenteerd, vergezeld van een aantal kernbegrippen om uit te leggen wat het levensgebied omvat. De 12 levensgebieden van de I.ROC zijn onderverdeeld in 4 hoofd domeinen (*home, opportunity, people* en *empowerment*; HOPE). Onder deze 4 domeinen vallen de volgende 12 levensgebieden:

- i. *home* (een veilige plek om te leven: 1. geestelijke gezondheid, 2. dagelijkse vaardigheden, 3. je veilig en op je gemak voelen);
- ii. *opportunity* (zinnvolle levensinvulling: 4. lichamelijke gezondheid, 5. bewegen en actief zijn, 6. zingeving en levensdoelen);
- iii. *people* (vrienden, vertrouwelingen en anderen die steun bieden: 7. zelfwaardering, 8. persoonlijk netwerk, 9. sociaal netwerk);
- iv. *empowerment* (regie voeren over eigen leven: 10. hoop voor de toekomst, 11. zelfmanagement, 12. participatie en controle).

Elk van de 12 items heeft 6 antwoordopties: (1 = nooit, 2 = bijna nooit, 3 = soms, 4 = regelmatig, 5 = bijna altijd, 6 = altijd). Zie [figuur 1 \(online\)](#).

De respondent (cliënt, zorgvrager of burger) wordt gevraagd om op een 6-puntsschaal een oordeel te geven over de afgelopen drie maanden. Bijvoorbeeld: 'In de afgelopen drie maanden voelde ik me geestelijk en emotioneel goed, gezond en gelukkig.' Een hogere score duidt op beter functioneren op de levensgebieden en meer welbevinden. Scores op de 12 items werden voor dit onderzoek opgeteld tot een totaalscore (uitersten: 12-72).

## Statistische analyse

### PR-scores

Percentielrangordescorres (PR-scores) worden direct berekend vanuit de ruwe scores en drukken de uitzonderlijkheid van een score uit op een schaal van 0 tot 100. Ze geven aan welk percentage van de referentiegroep een lagere score heeft.<sup>25,26</sup> De PR-score is gelijk aan het percentage van de oppervlakte onder de frequentieverdeling van scores links van de ruwe score (zie [figuur 2](#)). De PR-score wordt voor elke ruwe score berekend met de volgende formule:

$$PR = (\text{CumF}(\text{RS}) - 0,5 * F(\text{RS})) / N$$

Daarbij staat CumF(RS) voor de cumulatieve frequentie van de ruwe score en alle lagere scores, F(RS) staat voor de frequentie van de ruwe score zelf en N voor het totaal aantal scores in de steekproef.<sup>26</sup>

We demonstreren de formule aan de hand van een rekenvoorbeeld. Voor elke ruwe score wordt eerst de frequentie van voorkomen bepaald bij een normatieve referentiegroep. Zo hadden in ons I.ROC-databestand 114 deelnemers een ruwe score van 57 en 1325 deelnemers een score van 57 of lager. Deze frequenties worden omgezet naar proporties ( $n = 114$  van de 2366 deelnemers:  $F(57) = 4,8\%$ ) en de cumulatieve proportie ( $n = 1325$ ;  $\text{CumF}(57) = 56,0\%$ ). De cumulatieve proportie min de helft van de proportie die precies deze score heeft,  $56,0 - (4,8\%/2 = 2,4\%)$ , geeft de PR-score: 53,6 of afgerond naar een geheel getal  $PR = 54$ .

### T-scores

De T-score is al in 1922 door McCall voorgesteld als alternatief voor de gewone standaardscore of Z-score.<sup>27</sup> Deze standaardscore wordt berekend door van de ruwe score het populatiegemiddelde af te trekken en het resultaat te delen door de standaarddeviatie van de populatie. Dat geeft een score tussen grofweg -4 en 4 met een gemiddelde van 0. Elk punt op de schaal is een standaarddeviatie verwijderd van naastliggende punten (zie [figuur 2](#)). De T-score is ook een standaardscore en wordt in het eenvoudigste geval berekend door de Z-score met 10 te vermenigvuldigen en er 50 punten bij op te tellen. Zo ontstaat een schaal van grofweg 10 tot 90. Deze eenvoudige benadering geeft echter afwijkende waarden wanneer de ruwe scores niet normaal verdeeld zijn. In dat geval zijn gemiddelde en spreiding niet voldoende om de scoreverdeling te beschrijven.

Bij niet-normaal verdeelde ruwe scores is het zaak eerst de scores te normaliseren.<sup>15</sup> Dat kan met een benadering volgens de itemresponstheorie (IRT) of met een benadering die is gebaseerd op de frequentieverdeling van de scores (*rankit*). Omdat de IRT-benadering extra eisen stelt aan de psychometrische kenmerken van het instrument, zoals unidimensionaliteit, achtten we het beter om bij de I.ROC de rankitmethode toe te passen. Volgens de rankitmethode worden de ruwe scores op de I.ROC eerst omgezet naar PR-scores. Deze worden vervolgens omgezet naar Z- en T-scores. Wanneer op

deze wijze T-scores zijn verkregen, kan een oversteek-tabel van ruwe scores naar T-scores worden opgesteld. Voor de T-scoreschaal is 50 het middelpunt, 68% van alle mensen heeft een score tussen 40 en 60, 95% tussen 30 en 70 en bijna iedereen (99,7%) scoort tussen 20 en 80 (zie **figuur 2**).

**Figuur 2** illustreert hoe T- en PR-scores zich verhouden tot Z-scores indien die laatste normaal verdeeld zijn. Ook de I.ROC-schaal voor de ruwe totaalscore is in deze figuur opgenomen. **Figuur 2** laat zien dat rond de gemiddelde score van 50 de stappen op de PR-scoreschaal bijna een factor 4 groter zijn dan op de T-scoreschaal: T45-T55 = PR31-PR69. Verschillen in ruwe scores of T-scores worden rond dit gemiddelde met de PR-schaal als het ware uitvergroot. Aan de uiteinden van de schaal is het omgekeerde het geval en worden verschillen in T-scores samengedrukt met de PR-schaal. Het interval T70-T80 van 10 T-punten komt overeen met een interval van 2,2 PR-punt: PR97,7-PR99,9.

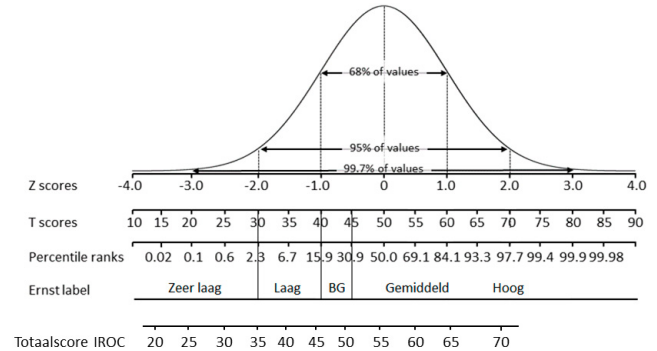
T-scores kunnen beschouwd worden als intervallschaal, PR-scores niet. PR-scores zijn volgordelijke waarden op een ordinale schaal. Een consequentie hiervan is dat vooruitgang richting herstel wel met een verschilscore als Delta T kan worden uitgedrukt, maar niet met een verschilscore op de PR-schaal. Een verschuiving van 10 punten op de PR-schaal heeft uiteenlopende betekenissen, afhankelijk van de positie op de schaal.

In dit onderzoek vergeleken we twee manieren van bepalen van T-scores: 1. T-scores volgens een lineaire formule op basis van de gemiddelde en standaarddeviatie van de ruwe scores en 2. genormaliseerde T-scores volgens een curvilineaire formule die we vonden met de rankitbenadering. De rankitmethode bestaat eruit dat in een databestand van een steekproef eerst voor alle ruwe scores PR-scores worden bepaald. T-scores kunnen vervolgens berekend worden door gebruik te maken van de vaste relatie tussen PR-scores en standaardcores. Zo kan een tabel gemaakt worden met ruwe scores, T- en PR-scores. Met curve-fitting kan de formule worden gevonden die de relatie tussen ruwe scores en T-scores (en tussen ruwe scores en PR-scores) optimaal benadert. Zo'n tabel of formule is nodig om de score van een individuele cliënt om te zetten naar een T-score. Een formule kan bijvoorbeeld ingebouwd worden in software die gebruikt wordt om de I.ROC af te nemen of te scoren en die de ruwe score, PR-score en T-score kan geven. Alle analyses voor dit onderzoek voerden we uit met R. Eerder publiceerden we een gedetailleerde beschrijving van de methode en de gebruikte R-code.<sup>15</sup>

## RESULTATEN

Uit de resultaten bleek dat ruwe totaalscores op de I.ROC al behoorlijk normaal verdeeld waren. De gemiddelde score op de I.ROC was 55,31 (SD: 8,76). De scheefheid (*skewness*) was -0,56; de gepiekttheid (*kurtosis*) was 0,28. Deze waarden vielen binnen de gewenste waarden

**Figuur 2. De normaalverdeling, Z-scores, T-scores (curvilineair), PR-scores, betekenisverlening in een label en de schaal voor de ruwe I.ROC-totaalscore**

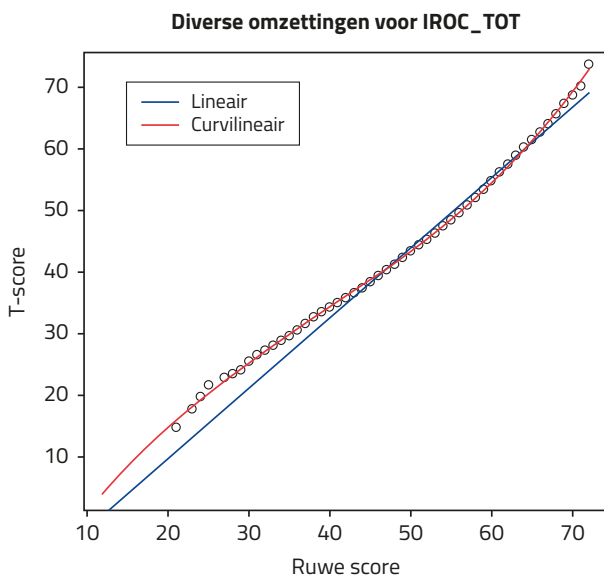


van -1 en 1, om uit te kunnen gaan van een normaalverdeling. De licht negatieve scheefheid was terug te voeren op een geringe overmaat aan extreem lage scores. De gemiddelde T-score na normaliseren was 49,98 (SD: 9,94; scheefheid: -0,04; gepiekttheid: -0,12); dit lag zeer nabij het ideaal van een normaalverdeling met gemiddeld 50 (SD: 10). Volgens een t-test (tweezijdig) was er geen significant sekseverschil in I.ROC-scores (p = 0,58): de gemiddelde score bij vrouwen (n = 1274) was 55,41 (SD: 8,65) en bij mannen (n = 1092) 55,20 (SD: 8,90). Tevens was er geen significant verband tussen leeftijd en totaalscore op de I.ROC (r = 0,00; p = 0,95) en ook niet tussen drie leeftijdsgroepen volgens een variantieanalyse (p = 0,503). De gemiddelde scores bij de leeftijdsgroepen waren: bij 18-46 jaar (n = 804): 55,10 (SD: 8,92); bij 47-65 jaar (n = 811): 55,59 (SD: 8,80); en bij 66 en ouder (n = 751): 55,22 (SD: 8,55).

**Figuur 3** laat de relatie zien van de ruwe scores op de x-as en T-scores op de y-as volgens een lineaire formule (T = 1,14\*RS-13,13), berekend op grond van gemiddelde en SD (de blauwe doorgetrokken lijn). Tevens zijn met bolletjes de T-scores weergegeven die volgens de rankitmethode werden bepaald (genormaliseerde T-scores). Ten slotte is met een rode lijn aangegeven welke T-scores de formule voor een curvilineaire relatie opleverde. De lineaire formule gaf op zich al een goede benadering van T-scores, maar in de lage scorereange leidde de lineaire formule tot een onderschatting van de T-score die wel tot 6,3 punten opliep (vergelijk bijvoorbeeld de lineaire met de rankitmethode bij een ruwe score van 25 in **tabel 1 (online)**). Ook bij extreem hoge scores was er een afwijking van 2 tot 3 T-scorepunten. De curvilineaire lijn benaderde de genormaliseerde (rankit-)T-scores beter.

De formule voor de relatie tussen ruwe scores en rankit-T-scores werd bepaald met curve-fitting als een derdegraads polynoom (zie de noot onder **tabel 1**). Deze formule kan gebruikt worden om in te bouwen in vragenlijstsoftware om de I.ROC te scoren. De T-scores

**Figuur 3. De bolletjes geven de relatie weer tussen ruwe scores (x-as) en op rankit gebaseerde T-scores (y-as); de lijnen geven de relatie volgens de lineaire en de curvilineaire formule**



die deze formule opleverde, zijn weergegeven in de kolom 'curvilineair' in **tabel 1**. Vanaf ruwe score 21 (de laagste score in deze steekproef) gaf deze curvilineaire formule een erg goede benadering van de met rankit gevonden T-scores, wat ook te zien is in **figuur 3**. Ten slotte werd ook met curve-fitting een formule bepaald voor de relatie tussen ruwe scores en PR-scores zoals die weergegeven zijn in **tabel 1 (online)**.

De drie kolommen met T-scores in **tabel 1** maken een gedetailleerde vergelijking mogelijk van hetgeen de drie benaderingen opleverden. De rankitbenadering gaf genormaliseerde T-scores op basis van de feitelijke frequentie van ruwe scores in de steekproef. Vergelijken met de rankit-T-scores gaf de lineaire benadering bij lage scores ( $RS < 32$ ) en bij de hoogste score ( $RS = 72$ ) een onderschatting van bijna 4 punten. De curvilineaire formule benaderde de rankit-T-scores beter (zie ook **figuur 2**). In **tabel 1** is tevens de PR-score weergegeven zoals gevonden in het databestand. Dit is het percentage respondenten met een lagere score dan de ruwe score plus de helft van het percentage met precies de ruwe score. De grens voor de laagste 5% lag bij een ruwe score van  $RS = 39$ . Bij de ruwe score 50 was de PR-score 25, een ruwe score van 56 correspondeerde met  $PR = 49$ , et cetera.

Als een hogere score meer gezondheid of herstel betekent, wordt als afgrenzing voor klinisch significante verandering (*clinical significant change*; CSC) op de T-scoreschaal vaak een  $T > 45$  gehanteerd (<https://www.healthmeasures.net/score-and-interpret/interpret-scores/promis/promis-score-cut-points>). Dat stemt overeen met een ruwe score van  $RS > 51$ . Ter vergelijking: de enige data over CSC (Engelse data van Rudd, gepubli-

ceerd door Baufeldt en Dawson in hun tabel 1) leveren als grenswaarde een ruwe score van minstens  $RS > 49$ , wat correspondeert met een grenswaarde  $T > 42$ .<sup>28</sup>

**Figuur 4 (online)** laat voor de I.ROC de relatie zien tussen ruwe scores, T- en PR-scores. Door de klokvorm van de normaalverdeling was er een sigmoïde (s-vormige) relatie tussen PR-scores en T-scores; de figuur illustreert dat de relatie tussen T- en PR-scores rond het gemiddelde lineair was ( $PR = 4 * T - 150$ ), maar dat bij extreem hoge of extreem lage T-scores de PR-scores dicht bij elkaar lagen. Er wordt in de figuur ook een betekenis gegeven aan de uitslagen in scoreklassen. Die verschilden enigszins voor T- en PR-scores (zie ook **tabel 1**).

## DISCUSSIE

De belangrijkste bevindingen van dit onderzoek zijn dat berekening volgens een lineaire vergelijking en met de rankitbenadering in het geval van de I.ROC elkaar niet veel ontlopen. Ze leiden alleen in de lage en in de hoge regionen van de schaal tot verschillende T-scores. De ruwe scores van de I.ROC waren al bij benadering normaal verdeeld en dan zijn het gemiddelde en de standaardafwijking goede uitgangspunten voor een lineaire omzetting en is er weinig verschil met de curvilineaire methode. We kennen echter ook voorbeelden van veelgebruikte meetinstrumenten, waarbij de lineaire benadering spaak loopt, zoals de *Minnesota Multiphasic Personality Inventory* (MMPI) of de *Brief Symptom Inventory* (BSI) en de *Vierdimensionele Klachtenlijst* (4DKL).<sup>15,29</sup> Hier biedt een curvilineaire omzetting uitkomst.

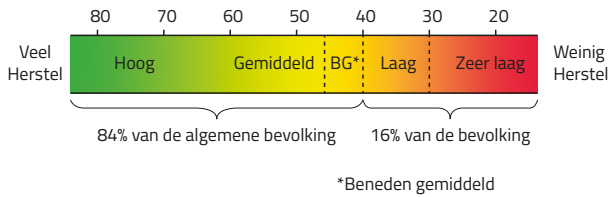
Bij de I.ROC volstaat voor de meeste situaties de toepassing van een eenvoudige formule om tot T-scores te komen:  $T = 1,14 * RS - 13,13$ . Deze formule kan heel eenvoudig 'met de hand' worden toegepast. Echter, bij lage ruwe scores ( $< 30$ ) geeft de lineaire formule een aanzienlijk lagere T-score dan de rankitbenadering (zie **figuur 3**) en voor inbouw in scoringssoftware is de ingewikkelde en preciezere curvilineaire formule te prefereren. Een op rankit gebaseerde T-score opzoeken in **tabel 1** of berekenen met een curvilineaire formule geeft het beste resultaat, want er wordt zo optimaal rekening gehouden waar ruwe scores op de I.ROC afwijken van een normale verdeling. Wie niet de T-score wil opzoeken of geen gebruik maakt van software om de I.ROC af te nemen kan met een simpele formule ( $T = 1,14 * RS - 13,13$ ) toch een T-score berekenen die voor de meeste ruwe I.ROC-scores voldoet.

### Nut van de T- en PR-scores

Een uitgebreid pleidooi voor T- en PR-scores is elders verwoord.<sup>17</sup> Kort samengevat, komt het erop neer dat omzetten van scores naar een universele schaal de communicatie over vragenlijstuitslagen tussen professionals onderling en tussen professionals en hun cliënten vereenvoudigt. Een algemene interpretatie van



**Figuur 5. Een algemene interpretatie van de T-scores voor functioneren**



Bij een herstelschaal als de I.ROC ligt de grens tussen ziek en gezond bij  $T < 45$  (ruwe score  $< 51$ ) en duidt een hogere score op meer herstel.

de T-scoreschaal voor scores waarbij een hoge score gezondheid weergeeft, is weergegeven in **figuur 5** (vgl. <https://www.healthmeasures.net/score-and-interpret/interpret-scores/promis/promis-score-cut-points>). De grensscore van  $T = 45$  geldt voor meer meetinstrumenten waarbij een hogere score voor meer herstel of meer welbevinden staat (zoals De Beurs e.a. eerder beschreven voor de *Mental Health Continuum-Short Form*; MHC-SF).<sup>30</sup> Het resultaat uit onze studie dat 29% van de bevolking op de I.ROC een benedengemiddelde T-score heeft ( $T < 45$ , corresponderend met een ruwe score  $RS < 52$ ), komt goed in de buurt van schattingen van de 12-maandsprevalentie van psychische stoornissen die bij het Trimbos-instituut (20-25%) uit hun NEMESIS-onderzoeken naar voren komen.<sup>31,32</sup> Dit ondersteunt de validiteit van de I.ROC als meetinstrument en de bruikbaarheid van de grensscore van  $T < 45$ .

### Sterke kanten en beperkingen

We beschikten over een ruime set aan gegevens, vanuit een representatieve steekproef van de algemene bevolking (LISS-panel), die bovendien door de onderzoekers geschoond werden van (te) snel ingevulde vragenlijsten. Lineaire en genormaliseerde T-scores werden op een zorgvuldige manier vastgesteld en formules voor de T- en de PR-score werden met geavanceerde statistiek (rankit, curve-fitting) bepaald. De resultaten brengen voor- en nadelen van de verschillende benaderingen om tot T-scores te komen aan het licht.

Voor de validiteit van normgegevens is het uiterst belangrijk dat de steekproef representatief was voor de Nederlandse bevolking. Huishoudens die niet over digitale middelen beschikten om vragenlijsten in te vullen, werden hierin vanuit CenterData Tilburg University gefaciliteerd. Ondanks deze inspanningen zijn mensen die geen Nederlands spreken en die minder digitale vaardigheden hebben waarschijnlijk ondervetwoordigd. Deze groep rapporteert over het algemeen een slechtere gezondheid, waardoor de gegevens van de steekproef mogelijk enigszins vertekend zijn met minder lage scores.<sup>33</sup>

Een andere beperking is dat we in dit artikel alleen normgegevens van de algemene populatie presenteren. Op dit moment worden in verscheidene ggz-instellingen

I.ROC-gegevens verzameld bij diverse cliëntengroepen. Dat zal ons in staat stellen in de nabije toekomst te onderzoeken of er verschillen zijn in (het profiel van) scores op de I.ROC in diverse ggz-settings. Als er betekenisvolle verschillen tussen diverse cliëntengroepen gevonden worden, dan worden daar binnenkort normgegevens voor beschikbaar gesteld.

## CONCLUSIE

Uit eerder onderzoek komt de I.ROC naar voren als een betrouwbaar, valide en bruikbaar meetinstrument. De meetpretentie van de I.ROC is breed en past goed bij de visie dat herstel op uiteenlopende aspecten van functioneren geïnventariseerd moet worden. T-scores zijn nuttig en zijn voor de I.ROC te bepalen met een eenvoudige formule. Voor inbouw in vragenlijstsoftware raden we echter toepassing van de complexere en preciezere formule aan. Ook zijn T- en PR-scores op te zoeken in de oversteektabel bij dit artikel (**tabel 1 online**). Gebruik van universele schalen wordt gepropageerd als een goede aanpak om uitvoering van ROM in de klinische praktijk te stimuleren, zowel om verschillende vragenlijsten naast elkaar te kunnen gebruiken als om de uitkomsten op een begrijpelijke manier te presenteren voor professionals én cliënten.

## LITERATUUR

- Slade M, Amering M, Farkas M, e.a. Uses and abuses of recovery: implementing recovery-oriented practices in mental health systems. *World Psychiatry* 2014; 13: 12-20.
- Davidson L, Roe D. Recovery from versus recovery in serious mental illness: One strategy for lessening confusion plaguing recovery. *J Ment Health* 2007; 16: 459-70.
- Leamy M, Bird V, Boutillier CL, e.a. Conceptual framework for personal recovery in mental health: systematic review and narrative synthesis. *Br J Psychiatry* 2011; 199: 445-52.
- Westerhof GJ, Keyes CL. Geestelijke gezondheid is meer dan de afwezigheid van geestelijke ziekte. *Maandblad Geestelijke Volksgezondheid* 2008; 63: 808-20.
- Huber M, Knottnerus JA, Green L, e.a. How should we define health? *BMJ* 2011; 343: 235-7.
- Huber M, van Vliet M, Boers I. Heroverweeg uw opvatting van het begrip 'gezondheid'. *Ned Tijdschr Geneesk* 2016; 160: A7720.
- Monger B, Hardie SM, Ion R, e.a. The individual recovery outcomes counter: preliminary validation of a personal recovery measure. *The Psychiatrist* 2013; 37: 221-7.
- Dickens GL, Rudd B, Hallett N, e.a. Factor validation and Rasch analysis of the individual recovery outcomes counter. *Disabil Rehabil* 2019; 41: 74-85.
- Rudd B, Karatzias T, Bradley A, e.a. Personally meaningful recovery in people with psychological trauma: Initial validity and reliability of the Individual Recovery Outcomes Counter (I. ROC). *Int J Ment Health Nurs* 2020; 29: 387-98.
- Beckers T, Koekkoek B, Hutschemaekers G, e.a. Measuring personal recovery in a low-intensity community mental healthcare setting: validation of the Dutch version of the Individual Recovery Outcomes Counter (I. ROC). *BMC Psychiatry* 2022; 22: 1-10.

De overige literatuurverwijzingen zijn online te raadplegen.