

# The bumpy road to achieve reliability of clinical profile characteristics in psychosis and related disorders

Steven Berendsen<sup>1,2</sup>  | Mirjam J. van Tricht<sup>1</sup> | Amy Tedja<sup>1</sup> | Thijs J. Burger<sup>1,2</sup> | Mariken B. de Koning<sup>2</sup> | Lieuwe de Haan<sup>1</sup>

<sup>1</sup>Department of Psychiatry, University Medical Center Amsterdam, Location Academic Medical Center, Amsterdam, The Netherlands

<sup>2</sup>Arkin Mental Health Care, Amsterdam, The Netherlands

## Correspondence

Steven Berendsen, Meibergdreef 9, 1105 AZ Amsterdam, Department of Psychiatry, Academic Medical Center, The Netherlands.  
Email: [s.berendsen@amsterdamumc.nl](mailto:s.berendsen@amsterdamumc.nl)

## Abstract

**Objectives:** Profile characteristics are factors that are relevant for diagnosis, prognosis or treatment. The present study aims to develop a set of clinically relevant profile characteristics. Moreover, our goal is to determine the inter-rater reliability (IRR) of the selected profile characteristics.

**Methods:** Potential profile characteristics were determined by literature review. Assessment of IRR was done by comparing scores on profile characteristics determined by two researchers. We conducted three subsequent studies: (1) assessment of pre-training IRR, (2) IRR following implementation of an instruction manual, (3) IRR after optimizing scoring methods. IRR was measured with the Intraclass Correlation Coefficient (ICC).

**Results:** IRR scores of profile characteristic Illegal activities were high across the three studies ( $ICC \geq 0.75$ ). Following training procedures in study 2 and 3, reliability estimates remained low to moderate ( $ICC < 0.75$ ) for the profile characteristics Support of relatives, Aggression recent and lifetime, substance use and insight recent. IRR scores of the other eight profile characteristics varied from low, moderate to high across studies.

**Conclusion:** IRR scores of profile characteristics were highly variable, and mostly inadequate in all three studies. Consequently, further research should focus on specification of severity scores of profile characteristics, optimizing scoring methods and re-evaluation of IRR.

## KEYWORDS

clinical profiling, heterogeneity, inter-rater reliability, psychosis, training procedures

## 1 | INTRODUCTION

The heterogeneity of schizophrenia and related disorders is omnipresent. Manifestations of the disease vary strongly in terms of symptomatology, vocational functioning and social participation (Kahn et al., 2015; Owen, Sawa, & Mortensen, 2016). Furthermore,

incidence and prevalence of psychiatric and somatic co-morbidity vary between patients, contributing to the heterogeneity of diagnosis, prognosis and treatment (Sprah, Dernovsek, Wahlbeck, & Haaramo, 2017; Yum, Hwang, Nasrallah, & Opler, 2016).

To capture this substantial variation, new methods are being developed to provide a more precise diagnosis, examples are network

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. International Journal of Methods in Psychiatric Research published by John Wiley & Sons Ltd.

analysis (Borsboom & Cramer, 2013), the trans-diagnostic approach (Nelson, McGorry, Wichers, Wigman, & Hartmann, 2017) or clinical profiling (Wigman et al., 2013). These novel methods aim to disentangle the diverse presentation and outcome parameters of the disease into specific categories, which are relatively easy to recognize and may form a target for intervention (Kahn & Keefe, 2013). Eventually, the ultimate goal is to develop diagnostic tools that can improve prediction of outcome, individualized treatment and prevention of illness progression.

Clinical profiling have been proposed to create a comprehensive set of clinical relevant prognostic factors (Wigman et al., 2013). Profile characteristics may represent current as well as life time patient characteristics, relevant for diagnosis and prognosis. By profiling we may personalize the diagnosis by targeting the large variation in symptomatology, comorbidity and social participation.

Clinical profiling of psychiatric illness may offer various benefits at the level of scientific research and clinical practice, in comparison to other methods that aim to address heterogeneity. First, a set of clinically relevant factors could support clinicians to ensure quick recognition of specific problem areas and specify their treatment-plan according to these profile characteristics. In fact, clinical profiling surpasses the current diagnostic categories and can be utilized in the trans-diagnostic approach. Second, profile characteristics can be easily applied by clinicians at individual patient-level, in comparison to the network approach which is predominantly applied on group-level symptoms in research (Bos & Wanders, 2016). To perform network analysis on individual patient level longitudinal time series of experienced sampling are needed (Yang et al., 2018). Third, profiling in scientific research could provide a more comprehensive and detailed description of patient characteristics, and thereby contribute to the selection and comparison of study populations. Last, for several of the proposed profile characteristics specific therapeutic interventions are available, such as interventions to enhance social interactions, health education or medication adherence training (Abdel Aziz et al., 2016; Dodell-Feder, Tully, & Hooker, 2015).

Until now a comprehensive overview of profile characteristics for psychosis and related disorders is missing. In addition, no studies have yet addressed the inter-rater reliability (IRR) of assessing profile characteristics. Assessment of IRR is an important first step in the psychometric evaluation, considering that sufficient reliability of profile characteristics is a pre-requisite for any further validity research (Kobak, Kane, Thase, & Nierenberg, 2007). Additionally, poor reliability leads to high variability of data and reduced power to find relevant differences (Perkins, Wyatt, & Bartko, 2000). An important tool to achieve sufficient reliability is training of raters, and this training should include information about profile characteristics, specifications for ratings and feedback on performance by experienced clinicians. Typically, training of raters leads to an increase in reliability estimates (Kobak, Engelhardt, Williams, & Lipsitz, 2004).

Consequently, in the present study we first aim to develop an overview of clinically relevant and feasible profile characteristics,

and subsequently determine the IRR of these profile characteristics in patients presenting with recent onset psychotic symptoms. Third, we set out to evaluate whether training procedures for raters improve IRR. We hypothesize that IRR scores of profile characteristics without training procedures will be insufficient, and training of raters will substantially improve reliability estimates.

## 2 | METHOD

### 2.1 | Sample characteristics

In Table 1 the demographical and clinical characteristics of the separate studies are shown. In total we included 139 patients, of whom we included 99 patients in study 1, 20 patients in study 2 and another 20 patients in study 3. We included a smaller sample in study 2 and 3 as we only aimed to evaluate the effect of training procedures on IRR. The majority of patients were diagnosed with a schizophrenia spectrum disorder and used antipsychotic medication. No significant differences in age, gender, global assessment of functioning (GAF), use of antipsychotic medication or diagnosis were found between the samples that participated in the three studies.

### 2.2 | Design

The design of the current study was cross-sectional and it contained three sub-studies. We included outpatients from the diagnostic center of the Early Psychosis Unit at the department of psychiatry, at the University Medical Center Amsterdam, location Academic Medical Center. Three independent clinicians (authors AT and TJB, IL in the acknowledgments) performed inter-rater reliability assessments across the sub-studies.

### 2.3 | Procedure

The main reason for referral to our diagnostic center was recent onset of psychotic symptoms. The diagnostic interview was conducted by a psychiatry resident or psychologist in training for healthcare psychologist, supervised by a clinical psychologist and psychiatrist. The psychiatrist and clinical psychologist were present during the first 30 min of the interview. After the diagnostic interview a comprehensive formal report was written by the psychiatry resident or psychologist in training, supervised by the psychiatrist or clinical psychologist. This report was written in a predefined format, and included the current psychiatric condition, illness history, information from relatives and other relevant information for diagnosis and treatment. It was a comprehensive document of approximately four to six pages. This formal report based on the diagnostic interview was used for IRR purposes.

TABLE 1 Demographic and clinical characteristics

	Study 1 (n = 99)	Study 2 (n = 20)	Study 3 (n = 20)	Between-groups	p-value
Gender (% male)	68.04%	55.00%	80.00%	$\chi^2$ (2.502)	0.286
Age (SD)	24.32 (6.0)	25.47 (5.14)	23.50 (3.92)	$F = 0.056$	0.946
GAF mean (SD)	52.28 (15.48)	50.65 (16.65)	50.00 (14.53)	$F = 0.147$	0.864
% Use of antipsychotic medication	86.67%	85.00%	73.33%	$\chi^2$ (0.724)	0.696
Diagnosis	-	-	-	$\chi^2$ (4.697)	0.320
% Schizophrenia spectrum disorders	84.69%	85.00%	73.33%	-	-
% Bipolar disorder	9.18%	10.00%	26.67%	-	-
% Other psychotic disorders	6.12%	5.00%	0.00%	-	-

Abbreviations: GAF, global assessment of functioning; SD, standard deviation.

## 2.4 | Profile characteristics

We selected clinically relevant profile characteristics based on an extensive review performed by the committee that developed the clinical guidelines for the diagnosis and treatment of psychotic disorders in the Netherlands (Trimbos, 2019). To substantiate the latter review we evaluated whether the selected profile characteristics in psychosis and related disorders were significantly associated with increase in psychotic symptoms or psychiatric re-hospitalization in follow-up studies (Alvarez-Jimenez et al., 2012; Braehler et al., 2013; de Haan, van Amelsvoort, Dingemans, & Linszen, 2007; Emsley, Chiliza, Asmal, & Harvey, 2013; Kooyman, Dean, Harvey, & Walsh, 2007; Michalska da Rocha, Rhodes, Vasilopoulou, & Hutton, 2018; Peter, Weiden, Grogg, & Locklear, 2004; Rabinovitch, Bechard-Evans, Schmitz, Jooper, & Malla, 2009).

Thereafter, our research group "Severe Psychotic Disorders of the University Medical Center Amsterdam of the University of Amsterdam and Arkin," including psychiatrists and psychologists with extensive clinical and scientific experience with psychotic disorders, provided input on clinical relevance of profile characteristics and on the feasibility to assess profile characteristics in regular practice. The selection of profile characteristics was founded on three aspects: (1) profile characteristics needed to be related to an increase in psychotic symptoms or psychiatric re-hospitalization, (2) profile characteristics should be easily applicable by clinicians and (3) profile characteristics should be clinically relevant for outcome measures in schizophrenia spectrum disorders. This resulted in the final set of profile characteristics that were used in our study.

## 2.5 | Inter-rater reliability

The general procedure of IRR assessment was as follows: two clinicians independently determined severity scores of profile characteristics by thoroughly evaluating the formal report written after the diagnostic assessment. The clinicians did not see the patient and only had access to the formal report. IRR was assessed by comparing severity scores of profile characteristics determined by both

clinicians. We chose this procedure because it is in line with clinical practice in which clinicians frequently base decisions on written information of other colleagues. Moreover it is in accordance with previous studies concerning reliability of medical diagnosis based on electronic record review (Chung, Chiang, Chou, Chu, & Chang, 2010; Kang et al., 2013; Varmdal et al., 2015). Estimation of IRR by multiple raters who merely observe video fragments may reduce information variance, leading to artificially inflated reliability estimates (Kobak et al., 2004). To evaluate the influence of training procedures on the IRR of profile characteristics, we conducted three subsequent studies:

### Study 1 (n = 99)

Pre-training IRR was determined by comparing severity scores of profile characteristics assessed by two independent researchers (A.T and I.L., mentioned in the acknowledgments). Both researchers had access to the description of the profile characteristics as shown in Table 2, there was no other training or instruction beforehand.

### Study 2 (n = 20)

Following findings of insufficient pre-training IRR in study 1, an instruction manual was developed in which specific criteria for each profile characteristic severity score were described with practical examples and instructions. Two researchers (A.T. and T.B.) thoroughly discussed the instruction manual and independently determined profile characteristics scores based on the formal report.

### Study 3 (n = 20)

Based on the findings of insufficient IRR of some profile characteristics after introducing an instruction manual and training we adjusted the instruction manual by extending criteria of severity scores. Furthermore, the designated sections in the formal report from which information should be retrieved to determine profile characteristics scores, were described. Profile characteristic specific terms were introduced to explore the report with the search-function, for example "school," "work" or "physical ill." With these adjustments we aimed to ensure that both raters retrieve the same information from the report and base their profile characteristic scores on the same data. Subsequently, two researchers (A.T. and T.B.) independently assessed profile characteristic scores based on a new set of formal reports.

TABLE 2 Description of profile characteristics

Profile characteristic	No/very mild problems	Moderate problems	Serious problems
Work or other activities	A paid job or equivalent of daily activities, e.g., volunteering, study, care taker tasks	Some structural activities, for instance work, volunteering, study/school, sports, social activities or care taker tasks for 2 days per week	Less than 2 days per week with structural activities
Support system	Adequate support by family/friends/partner, no or minimal problems	Moderate support or moderate problems, for instance dedicated support available only by 1 person or when support leads to infrequent conflicts	No or minimal support or serious problems or conflicts
Physical health	No or minimal physical health issues	Moderate physical health problems for instance being at risk for physical health problems, for instance metabolic complications of antipsychotic medicine or physical illness with moderate impact on functioning, for instance joint problems or well-regulated diabetes	Serious physical health problems with impact on daily life, for instance diabetes that is difficult to regulate
Substance usage (excluding nicotine) Lifetime Last month	None or limited use without consequences	Moderate use with moderate consequences for instance sporadically missing job or social appointments because of substance abuse	Serious abuse with serious consequences, for instance not being able to maintain a paid job, not being able to go to school, physical illness or financial problems because of substance abuse
Aggression Lifetime Last month	None	Verbal aggression or minimal physical aggression without physical harm towards self or others for instance damaging goods or personal belongings	Serious threats and/or physical aggression with physical harm towards self or others
Suicidality Lifetime Last month	No suicidal thoughts ever (thoughts with intent to end life)	Suicidal thoughts present, without suicide plan, no history of attempted suicide or suicidal actions	Suicide plan, history of attempted suicide, suicidal actions
Compliance Lifetime Last month	None or minimal problems	Moderate compliance, for instance infrequent irregular use of medication	Serious problems, for instance severe avoidance of mental health care
Adverse life events	None or minimal adverse life events	Adverse life events without serious physical or emotional threats, violence and/or aggression	Traumatic life events with serious physical, emotional threats, violence and/or aggression
Involuntary treatment	None	One episode of short term involuntary treatment	More than 1 episode of short term involuntary treatment or 1 episode of long term involuntary treatment
Illegal activities	None or minimal, for instance traffic fines	Moderate illegal activities, no imprisonment, no violence for instance stealing a bike	Serious crimes with imprisonment and/or violence
Illness insight Lifetime Last month	Good insight	Moderate insight, for instance recognizing Symptoms, but not psychiatric diagnosis	Minimal to no insight

## 2.6 | Ethic approval

Data used in the current study were anonymized by an administrative employee who was involved in treatment planning. Both researchers who assessed IRR based on formal reports had no access to non-anonymized data. Furthermore, the Dutch Central Medical Ethical Committee has ruled that Dutch Law regarding research with humans does not apply to the collection of anonymized information and, consequently, analyzing anonymized data for the present study does not require additional informed consent from participants.

## 2.7 | Statistical analysis

To determine the IRR between the two independent researchers we chose the Intraclass Correlation Coefficient (ICC, 2-way, mixed-effects model with absolute agreement) with 95% confidence interval. The ICC was chosen as IRR coefficient as our primary endpoint was an ordinal variable. Further, recent literature proposed more strict cut-off values for interpreting ICC values. Accordingly, we used the following ICC cut-off values: ICC <0.5 reflects poor agreement, ICC 0.5–0.75 reflects moderate agreement and ICC >0.75 reflects

high agreement (Koo & Li, 2016). Analysis of variance or Chi-Square tests were used to test for differences in gender, age, GAF, use of antipsychotic medication (yes or no) and diagnosis between studies.

### 3 | RESULTS

#### 3.1 | Profile characteristics

A detailed description of the profile characteristics and specifications of severity scores are shown in Table 2. The following profile characteristics were selected: substance use, aggression, suicidality, compliance, insight (lifetime prevalence and prevalence during last month was assessed), work or other activities, support of relatives, physical health, adverse life events, involuntary treatment and illegal activities (prevalence during last month was assessed). Profile characteristics were rated on a 3-point severity scale: (1) light problems or no problem, (2) moderate problems and (3) severe problems. A score of 0 was assigned in case of missing information.

#### 3.2 | Inter-rater reliability

In Table 3 IRR scores of profile characteristics in all studies are shown. In study 1, pre-training reliability estimates of profile characteristics Involuntary care and Illegal activities were high (ICC: 0.912 and 0.857). IRR scores of the majority of profile characteristics was moderate (ICC: 0.5–0.75), while IRR score of profile characteristics support of relatives and insight recent was low (ICC score: 0.410 and 0.397).

In study 2, IRR scores of profile characteristics life events, daily activities, substance use lifetime and illegal activities was high (ICC: 0.762–1.000), whereas most profile characteristics scored moderate reliability (ICC: 0.5–0.75). IRR scores of profile characteristics insight recent, aggression lifetime and compliance recent was low (ICC 0.391–0.424).

In study 3, IRR scores of the profile characteristics Physical Health, Suicidality lifetime, Suicidality recent, Compliance recent, Involuntary care and Illegal activities were high (ICC: 0.758–1.000). Reliability estimates of profile characteristics Substance use lifetime, Insight recent and Aggression recent was moderate (ICC: 0.555–0.655), while IRR scores of remaining four profile characteristics were low.

### 4 | DISCUSSION

In the present study we created comprehensive set of prognostic factors that were considered clinically relevant and feasible in regular clinical practice. We investigated the IRR of these profile characteristics in patients with recent onset psychotic disorders. In addition, we explored whether training procedures and scoring instructions improved reliability estimates of profile characteristics.

The selected profile characteristics represent important aspects of symptomatology, comorbidity and psycho-social functioning relevant for prognosis and treatment in psychosis and related disorders. Our findings showed that the majority of IRR scores of profile characteristics were insufficient, and that training procedures provided some improvement in IRR of some profile characteristics. However, after training, some profile characteristics actually showed reduced IRR.

We found no comparable studies concerning the IRR of profile characteristics, therefore we are not able to compare our results with previous studies. On the other hand, psychiatric classifications such as schizoaffective disorder or attention deficit disorder also show highly variable reliability estimates (Grabemann et al., 2017; Santelmann, Franklin, Busshoff, & Baethge, 2016). Even beyond the scope of psychiatry, the diagnosis of allergic laryngitis or classification systems for scoring tibia plateau fractures demonstrated insufficient or highly diverse IRR scores (Millar, Arnold, Thewlis, Fraysse, & Solomon, 2018; Stachler & Dworkin-Valenti, 2017). We therefore conclude that IRR of profile characteristics, psychiatric classifications and at least certain diagnoses in general medicine are highly variable.

When examining our results in more detail, pre-training IRR of profile characteristics in study 1 was generally inadequate, except for the profile characteristics Involuntary Care and Illegal Activities. These results are in line with various research concerning insufficient pre-training IRR of observational instruments (Berendsen et al., 2019; Muller & Dragicevic, 2003; Muller et al., 1998). Utilizing profile characteristics or other psychiatric instruments without instruction and training for raters should therefore be avoided.

Following implementation of the instruction manual four profile characteristics reached sufficient reliability, while IRR scores of the majority of profile characteristics remained poor. After the manual was complemented by optimizing scoring methods, and adding more detailed specifications and instructions for the raters, several profile characteristics achieved sufficient IRR scores, whereas the observed IRR score of eight profile characteristics remained insufficient.

Previous studies concerning training procedures followed by reliability assessment generally show improvement of IRR. For instance: a commonly applied instrument such as the Positive and Negative Syndrome Scale demonstrated a substantial increase in IRR after training (Kobak et al., 2004; Muller et al., 1998). On the other hand, there is conflicting evidence whether intensive training procedures for the Hamilton Depression Rating Scale (HDRS) actually lead to improvement of the IRR (Demitrack, Faries, Herrera, DeBrotta, & Potter, 1998; Muller & Dragicevic, 2003). Various reasons have been stated for the non-improving reliability estimates of the HDRS after training, for instance rater selection or shortcomings in item-content.

Despite great effort we were not able to achieve sufficient reliability scores for all profile characteristics, we hypothesize that this could be due to two main reasons. To start, although we aimed to set clear specifications of severity scores in the instruction manual, some

Profile characteristic	Study 1 (n = 99)	Study 2 (n = 20)	Study 3 (n = 20)
Involuntary care	<b>0.912 (0.872–0.940)</b>	0.653 (0.272–0.855)	<b>0.968 (0.923–0.987)</b>
Illegal activities	<b>0.857 (0.794–0.902)</b>	<b>1.000</b>	<b>1.000</b>
Suicidality lifetime	0.743 (0.640–0.820)	0.550 (–0.129–0.868)	<b>0.916 (0.746–0.975)</b>
Substance use lifetime	0.696 (0.578–0.786)	<b>0.897 (0.740–0.961)</b>	0.555 (0.017–0.820)
Aggression lifetime	0.644 (0.456–0.766)	0.400 (–0.159–0.809)	–0.103 (–0.429–0.323)
Daily activities	0.622 (0.484–0.729)	<b>0.762 (0.494–0.899)</b>	0.317 (–0.102–0.661)
Life events	0.579 (0.432–0.696)	<b>0.777 (0.456–0.919)</b>	0.231 (–0.142–0.610)
Compliance recent	0.575 (0.423–0.694)	0.391 (–0.025–0.704)	<b>0.758 (0.480–0.897)</b>
Physical health	0.548 (0.394–0.672)	0.571 (0.152–0.815)	<b>0.778 (0.517–0.906)</b>
Support of relatives	0.410 (0.232–0.562)	0.554 (0.143–0.798)	0.066 (–0.322–0.463)
Insight recent	0.397 (0.217–0.551)	0.424 (0.015–0.722)	0.596 (0.125–0.831)
Aggression recent	-	0.706 (0.216–0.911)	0.655 (0.317–0.846)
Suicidality recent	-	0.596 (0.233–0.816)	<b>0.764 (0.482–0.904)</b>
Substance use recent	-	0.678 (0.349–0.859)	0.422 (–0.028–0.725)

TABLE 3 IRR of profile characteristics

Abbreviation: IRR, inter-rater reliability. In bold ICC scores > 0.75.

profile characteristics might have been insufficiently defined to reach agreement between professionals. For instance, the description of severity scores of the profile characteristic support of relatives may have introduced too much room for interpretation. To illustrate: differentiating between moderate and severe scores, raters had to distinguish infrequent support of relatives (moderate) or serious problems with relatives (severe). The same difficulties in setting distinct boundaries between severity scores were found for the profile characteristics Aggression, Substance abuse recent or Insight recent. Identical concerns were found in achieving sufficient IRR for the HDRS, the authors suggested that limitations in definition of items of the HDRS may have resulted in poor reliability estimates (Demitrack et al., 1998).

Another possible cause of the observed low IRR might be that in the diagnostic report information was missing too frequently to adequately determine profile characteristic scores. Raters might have been obliged to select options based on lacking information, this could have led to an increased information variance which is a major source of low reliability. Nevertheless, our finding of large disagreement among raters when obtaining information from medical records is disturbing, considering the substantial number of studies that are based on collecting information from medical records.

An important question that remains is whether profiling is the best approach to address heterogeneity in psychiatric illness. Profile characteristics are easily applicable for clinicians and not labor intensive. However, it proved difficult to achieve overall reliability in the assessment of the profile characteristics. From that perspective, introduction into clinical practice or scientific research is not warranted at the current time, since insufficient reliability affects study power, placebo response and possible clinical decision making

(Berendsen et al., 2019; Kobak et al., 2010; Perkins et al., 2000). Consequently, effort should be focused on improving reliability of the assessment of these prognostic factors.

Our study should be viewed in light of several limitations. First, our sample size in study 2 and 3 is underpowered to permit strong conclusions concerning improvement or deterioration of reliability scores. Therefore, we conducted no formal statistical comparisons between ICC scores of the different studies. We merely tested improvement of ICC scores in the established categories for interpretation of IRR scores as low, moderate to high (Koo & Li, 2016). Second, we only tested IRR. Evaluating other forms of reliability, such as internal consistency or test-retest agreement is necessary to complement knowledge concerning the psychometric properties of profile characteristics. Third, we did not conduct a systematic search, therefore it is possible that not all studies with prognostic factors for patients with psychotic disorders and related illness were included.

We consider it a strength of our study that we created a comprehensive set of clinically relevant profile characteristics. Another advantage of our study are the training procedures for raters with repeated reliability assessment.

We found that IRR scores of clinical profile characteristics were highly variable and mostly poor, even after training procedures. Therefore, we strongly recommend to further develop definitions and trainings procedures of profile characteristics and evaluate IRR and other psychometric properties before clinical profiling can be introduced in clinical practice. Furthermore, we recommend that future research may explore the relationship between profile characteristics and underlying psychological mechanisms, for instance how the profile characteristic support system may be related to attachment disruptions (Fraley, 2019). Understanding these fundamental

mechanisms may support the search for novel targets of therapeutic intervention.

## ACKNOWLEDGMENTS

The authors thank Ijsbrand Leertouwer for his help with the data-collection. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

## CONFLICT OF INTERESTS

All authors declare not to have any conflicts of interest that might be interpreted as influencing the content of the manuscript.

## AUTHOR CONTRIBUTIONS

All authors contributed to the study design and proposal, literature search, analysis and interpretation. Authors Amy Tedja and Thijs J. Burger performed the data-collection, author Steven Berendsen drafted the manuscript and all other authors provided critical revisions. Authors Mirjam J. van Tricht, Mariken B. de Koning and Lieuwe de Haan supervised statistical analysis, study design and writing of the manuscript. All authors contributed to and have approved the final manuscript.

## ORCID

Steven Berendsen  <https://orcid.org/0000-0003-4080-5009>

## REFERENCES

- Abdel Aziz, K., Elamin, M. H., El-Saadouni, N. M., El-Gabry, D. A., Barakat, M., Alhayyas, F., & Moselhy, H. F. (2016). Schizophrenia: Impact of psychopathology, faith healers and psycho-education on adherence to medications. *International Journal of Social Psychiatry*, 62(8), 719–725. <https://doi.org/10.1177/0020764016676215>
- Alvarez-Jimenez, M., Priede, A., Hetrick, S. E., Bendall, S., Killackey, E., Parker, A. G., ... Gleeson, J. F. (2012). Risk factors for relapse following treatment for first episode psychosis: A systematic review and meta-analysis of longitudinal studies. *Schizophrenia Research*, 139(1–3), 116–128. <https://doi.org/10.1016/j.schres.2012.05.007>
- Berendsen, S., Kapitein, P., Schirmbeck, F., van Tricht, M. J., McGuire, P., Morgan, C., ... de Haan, L. (2019). Pre-training inter-rater reliability of clinical instruments in an international psychosis research project. *Accepted for publication in Schizophrenia Research*.
- Borsboom, D., & Cramer, A. O. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, 9, 91–121. <https://doi.org/10.1146/annurev-clinpsy-050212-185608>
- Bos, E. H., & Wanders, R. B. (2016). Group-level symptom networks in depression. *JAMA Psychiatry*, 73(4), 411. <https://doi.org/10.1001/jamapsychiatry.2015.3103>
- Braehler, C., Valiquette, L., Holowka, D., Malla, A. K., Joober, R., Ciampi, A., ... King, S. (2013). Childhood trauma and dissociation in first-episode psychosis, chronic schizophrenia and community controls. *Psychiatry Research*, 210(1), 36–42. <https://doi.org/10.1016/j.psychres.2013.05.033>
- Chung, M. H., Chiang, I. J., Chou, K. R., Chu, H., & Chang, H. J. (2010). Inter-rater and intra-rater reliability of nursing process records for patients with schizophrenia. *Journal of Clinical Nursing*, 19(21–22), 3023–3030. <https://doi.org/10.1111/j.1365-2702.2010.03301.x>
- Demitrack, M. A., Faries, D., Herrera, J. M., DeBroda, D., & Potter, W. Z. (1998). The problem of measurement error in multisite clinical trials. *Psychopharmacology Bulletin*, 34(1), 19–24.
- Dodell-Feder, D., Tully, L. M., & Hooker, C. I. (2015). Social impairment in schizophrenia: New approaches for treating a persistent problem. *Current Opinion in Psychiatry*, 28(3), 236–242. <https://doi.org/10.1097/ycp.000000000000154>
- Emsley, R., Chiliza, B., Asmal, L., & Harvey, B. H. (2013). The nature of relapse in schizophrenia. *BMC Psychiatry*, 13, 50. <https://doi.org/10.1186/1471-244X-13-50>
- Fraley, R. C. (2019). Attachment in adulthood: Recent developments, emerging debates, and future directions. *Annual Review of Psychology*, 70, 401–422. <https://doi.org/10.1146/annurev-psych-010418-102813>
- Grabemann, M., Zimmermann, M., Strunz, L., Ebbert-Grabemann, M., Scherbaum, N., Kis, B., & Mette, C. (2017). New ways of diagnosing ADHD in adults. *Psychiatrische Praxis*, 44(4), 221–227. <https://doi.org/10.1055/s-0043-100714>
- de Haan, L., van Amelsvoort, T., Dingemans, P., & Linszen, D. (2007). Risk factors for medication non-adherence in patients with first episode schizophrenia and related disorders; a prospective five years follow-up. *Pharmacopsychiatry*, 40(6), 264–268. <https://doi.org/10.1055/s-2007-992141>
- Kahn, R. S., & Keefe, R. S. (2013). Schizophrenia is a cognitive illness: Time for a change in focus. *JAMA Psychiatry*, 70(10), 1107–1112. <https://doi.org/10.1001/jamapsychiatry.2013.155>
- Kahn, R. S., Sommer, I. E., Murray, R. M., Meyer-Lindenberg, A., Weinberger, D. R., Cannon, T. D., ... Insel, T. R. (2015). Schizophrenia. *Nature Reviews. Disease Primers*, 1, 15067. <https://doi.org/10.1038/nrdp.2015.67>
- Kang, B. S., Cheong, H. K., Jung, K. Y., Jang, S. H., Yoo, J. K., Kim, D. W., ... Lee, S. Y. (2013). The validity and reliability of characterizing epilepsy based on an external review of medical records. *Epidemiology and Health*, 35, e2013006. <https://doi.org/10.4178/epih/e2013006>
- Kobak, K. A., Engelhardt, N., Williams, J. B., & Lipsitz, J. D. (2004). Rater training in multicenter clinical trials: Issues and recommendations. *Journal of Clinical Psychopharmacology*, 24(2), 113–117.
- Kobak, K. A., Kane, J. M., Thase, M. E., & Nierenberg, A. A. (2007). Why do clinical trials fail? The problem of measurement error in clinical trials: Time to test new paradigms? *Journal of Clinical Psychopharmacology*, 27(1), 1–5. <https://doi.org/10.1097/JCP.0b013e31802eb4b7>
- Kobak, K. A., Leuchter, A., DeBroda, D., Engelhardt, N., Williams, J. B., Cook, I. A., ... Alpert, J. (2010). Site versus centralized raters in a clinical depression trial: Impact on patient selection and placebo response. *Journal of Clinical Psychopharmacology*, 30(2), 193–197.
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Kooyman, I., Dean, K., Harvey, S., & Walsh, E. (2007). Outcomes of public concern in schizophrenia. *The British Journal of Psychiatry*, 191(S50), s29–s36. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/18019041>
- Michalska da Rocha, B., Rhodes, S., Vasilopoulou, E., & Hutton, P. (2018). Loneliness in psychosis: A meta-analytical review. *Schizophrenia Bulletin*, 44(1), 114–125. <https://doi.org/10.1093/schbul/sbx036>
- Millar, S. C., Arnold, J. B., Thewlis, D., Frayssé, F., & Solomon, L. B. (2018). A systematic literature review of tibial plateau fractures: What classifications are used and how reliable and useful are they? *Injury*, 49(3), 473–490. <https://doi.org/10.1016/j.injury.2018.01.025>
- Muller, M. J., & Dragicevic, A. (2003). Standardized rater training for the Hamilton Depression Rating Scale (HAMD-17) in psychiatric novices. *Journal of Affective Disorders*, 77(1), 65–69.
- Muller, M. J., Rossbach, W., Dannigkeit, P., Muller-Siecheneder, F., Szegedi, A., & Wetzel, H. (1998). Evaluation of standardized rater training for the positive and negative Syndrome scale (PANSS). *Schizophrenia Research*, 32(3), 151–160.

- Nelson, B., McGorry, P. D., Wichers, M., Wigman, J. T. W., & Hartmann, J. A. (2017). Moving from static to dynamic models of the onset of mental disorder: A review. *JAMA Psychiatry*, *74*(5), 528–534. <https://doi.org/10.1001/jamapsychiatry.2017.0001>
- Owen, M. J., Sawa, A., & Mortensen, P. B. (2016). Schizophrenia. *Lancet*, *388*(10039), 86–97. [https://doi.org/10.1016/S0140-6736\(15\)01121-6](https://doi.org/10.1016/S0140-6736(15)01121-6)
- Perkins, D. O., Wyatt, R. J., & Bartko, J. J. (2000). Penny-wise and pound-foolish: The impact of measurement error on sample size requirements in clinical trials. *Biological Psychiatry*, *47*(8), 762–766.
- Peter, J., Weiden, C. K., Grogg, A., & Locklear, J. (2004). Partial compliance and risk of rehospitalization among California medicaid patients with schizophrenia. *Psychiatric Services*, *55*(8), 886–891. <https://doi.org/10.1176/appi.ps.55.8.886>
- Rabinovitch, M., Bechard-Evans, L., Schmitz, N., Joobor, R., & Malla, A. (2009). Early predictors of nonadherence to antipsychotic therapy in first-episode psychosis. *Canadian Journal of Psychiatry*, *54*(1), 28–35. <https://doi.org/10.1177/070674370905400106>
- Santelmann, H., Franklin, J., Busshoff, J., & Baethge, C. (2016). Interrater reliability of schizoaffective disorder compared with schizophrenia, bipolar disorder, and unipolar depression—a systematic review and meta-analysis. *Schizophrenia Research*, *176*(2–3), 357–363. <https://doi.org/10.1016/j.schres.2016.07.012>
- Sprah, L., Dernovsek, M. Z., Wahlbeck, K., & Haaramo, P. (2017). Psychiatric readmissions and their association with physical comorbidity: A systematic literature review. *BMC Psychiatry*, *17*(1), 2. <https://doi.org/10.1186/s12888-016-1172-3>
- Stachler, R. J., & Dworkin-Valenti, J. P. (2017). Allergic laryngitis: Unraveling the myths. *Current Opinion in Otolaryngology & Head and Neck Surgery*, *25*(3), 242–246. <https://doi.org/10.1097/moo.0000000000000354>
- Trimbos. (2019). *Submitted for publication to early intervention in psychiatry*.
- Varmdal, T., Ellekjaer, H., Fjaertoft, H., Indredavik, B., Lydersen, S., & Bonna, K. H. (2015). Inter-rater reliability of a national acute stroke register. *BMC Research Notes*, *8*, 584. <https://doi.org/10.1186/s13104-015-1556-3>
- Wigman, J. T., van Os, J., Thiery, E., Derom, C., Collip, D., Jacobs, N., & Wichers, M. (2013). Psychiatric diagnosis revisited: Towards a system of staging and profiling combining nomothetic and idiographic parameters of momentary mental states. *PLoS One*, *8*(3), e59559. <https://doi.org/10.1371/journal.pone.0059559>
- Yang, X., Ram, N., Gest, S. D., Lydon-Staley, D. M., Conroy, D. E., Pincus, A. L., & Molenaar, P. C. M. (2018). Socioemotional dynamics of emotion regulation and depressive symptoms: A person-specific network approach. *Complexity*, *2018*. <https://doi.org/10.1155/2018/5094179>
- Yum, S. Y., Hwang, M. Y., Nasrallah, H. A., & Opler, L. A. (2016). Transcending psychosis: The complexity of comorbidity in schizophrenia. *Psychiatric Clinics of North America*, *39*(2), 267–274. <https://doi.org/10.1016/j.psc.2016.01.007>

**How to cite this article:** Berendsen S, van Tricht MJ, Tedja A, Burger TJ, de Koning MB, de Haan L. The bumpy road to achieve reliability of clinical profile characteristics in psychosis and related disorders. *Int J Methods Psychiatr Res*. 2021;30:e1858. <https://doi.org/10.1002/mpr.1858>