

## Clinicians have several therapeutic relationships and patients only one: the effect on their assessments of relationships

Article (Accepted Version)

Greenberg, Lauren, Bremner, Stephen, Carr, Catherine and Priebe, Stefan (2018) Clinicians have several therapeutic relationships and patients only one: the effect on their assessments of relationships. *International Journal of Methods in Psychiatric Research*, 27 (4). e1722 1-7. ISSN 1049-8931

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/77779/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

### **Copyright and reuse:**

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Clinicians have several therapeutic relationships and patients only one: the effect on their assessments of relationships

Lauren Greenberg<sup>a</sup>, Stephen Bremner<sup>b</sup>, Catherine Carr<sup>c</sup>, Stefan Priebe<sup>c</sup>

<sup>a</sup>Pragmatic Clinical Trials Unit, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, E1 2AB

<sup>b</sup>Department of Primary Care and Public Health, Brighton and Sussex Medical School, University of Brighton, Brighton, BN1 9PH

<sup>c</sup>Unit for Social and Community Psychiatry, WHO Collaborative Centre for Mental Health Services Development, Queen Mary University of London, Newham Centre for Mental Health, London E13 8SP

Corresponding author: Stefan Priebe (email: [s.priebe@qmul.ac.uk](mailto:s.priebe@qmul.ac.uk))

### **Acknowledgement**

We thank Prof Rose McCabe, University of Exeter, for allowing us to use data from her study of communication between patients and their psychiatrists.

## **Abstract**

### Objectives

Little attention has been given to the common assessment problem that clinicians assess outcomes of several patients and may rate them in comparison to one another, whereas patients assess only their own outcomes without any comparison. We explored empirically whether this would lead to a greater variability of clinician ratings as compared to patient ratings.

### Methods

Data from two independent samples in which clinicians and patients, using consistent instruments, rated their therapeutic relationships. We present descriptive statistics of variability and intraclass correlation coefficients (ICC).

### Results

The Helping Alliance Scale (HAS) was completed at baseline and follow-up by 20 clinicians and 103 patients in an observational study and by 88 clinicians and 431 patients in a trial. Patients tended to rate their relationship 5-10% more highly than their clinicians, but with 50-100% more variability. Intra-clinician HAS ratings were more correlated than those by patients (ICC 0.3-0.7 versus 0.0-0.2).

### Conclusions

Contrary to our assumption, clinicians' ratings of therapeutic relationships were in both samples less variable than those of their patients. When clinicians rate outcomes

of several patients, a cluster effect of ratings may have to be considered in the design and analysis.

**Keywords:** Helping Alliance Scale, therapeutic relationship, intracluster correlation coefficient, variability

## **Main text**

### **1. Introduction**

It is commonly known that in studies where the same clinician treats several patients, there is a clustering effect by clinician. The design and analysis of these studies must take account of this type of clustering (Kahan & Morris, 2013). However, little attention has been paid to the problem that in these studies the outcomes of several patients are rated by the same clinician, whereas patients rate themselves or – in the case of assessments of the therapeutic relationship - their clinician only. This is a distinctive type of clustering concerning the measurement of outcomes rather than the delivery of treatment. The situation explored in this paper is that of patients and clinicians assessing their therapeutic relationship. Therapeutic relationships are formed between a patient and a clinician and can be assessed by both of them. Whilst measurement of the therapeutic relationship is commonly conceptualised as a process variable, it offers the opportunity to compare the variability of clinician and patient ratings, as both rate the same construct, i.e. the relationship that exists between them. For the purposes of this paper, we use the term ‘outcome’ in the statistical sense, i.e. the independent variable in the model, as opposed to the therapeutic relationships as clinical outcome.

We consider two theories for how such a cluster effect of ratings may influence the ratings given. First, there may be a cluster effect of rating bias due to personality, experience, case mix, rapport or a tendency to view the therapeutic relationship through a “theoretical lens” (Baldwin, Wampold & Imel, 2007; Horvath, 2000). In line with theories regarding rater bias, clinicians may generally view their relationships more positively (‘halo effect’) or negatively (‘severity’), seek to be consistent in their responses (‘consistency effect’) or tend to avoid extremes (‘central tendency’) (Podsakoff, MacKenzie & Lee, 2003); The ratings of one clinician may tend to be significantly higher or lower than average, they may try to avoid extremes altogether or spread their ratings across the range of the scale. Some items may invite acquiescence or social desirability responses. For example, a clinician may feel obliged to demonstrate good relationships across all of their patients.

A second theory is that the variance of clinician ratings would be larger than the variance of the patient ratings from the same group. When clinicians rate several patients, they might compare patients and therefore increase the variance of ratings based on their experience of different therapeutic relationships ('contrast effect'). Conversely, patients rate only themselves or their clinician and therefore usually have no comparator for their assessment. Patients tend to rate their therapeutic relationships very positively (Ardito & Rabellino, 2011), and it is possible that five patients of one clinician all rate towards the positive extreme. A clinician might hesitate to rate the same five of their patients at the positive extreme, and feel that some distinction is required or rate on factors not related to the alliance, such as treatment adherence (Phillips, Leventhal & Leventhal, 2011; Thompson, Howell & McCabe, 2015). They might conclude that relationships with some patients, although still positive, are less so than those with others and should therefore be reflected in a lower rating, leading to a larger variance.

The consequences of extreme or central tendency ratings are to reduce the variability of responses, which may in result under or overestimation of the construct concerned (Streiner & Norman, 2008). However, in addition to this, clustering of ratings made by a single rater would violate the assumption of independence of ratings, which is a core assumption of simple statistical methods (Crits-Christoph & Mintz, 1991; Elvins & Green, 2008; Roberts, 1999). At the planning stage, ignoring this lack of independence in the sample size calculation is likely to lead to an underpowered study (Walwyn & Roberts, 2010). The inferential consequences of ignoring such clustering at the analysis stage are an increased risk of rejecting null hypotheses that are true, i.e. an increased type I error rate, and constructing confidence intervals that are too narrow (Lee & Thompson, 2005).

Studies that have so far examined the question of therapist effects, have often done so through the lens of the relative contribution to clinical outcomes or the apportioning of the variability of patient ratings within and between therapists (Baldwin et al., 2007; Del Re, Flückiger, Horvath, Symonds & Wampold, 2012; Dinger,

Strack, Leichsenring, Wilmers & Schauenburg, 2008; Imel, Hubbard, Rutter & Simon, 2013; Zuroff, Kelly, Leybman, Blatt & Wampold, 2010). These studies provide varying estimates of variability within-therapists (ie. ratings by patients on one therapist) and between-therapists (degree of clustering of patient ratings across therapists). Ratings at the between-therapist level tend to be lower, but a strong predictor of clinical outcomes (Baldwin et al., 2007; Del Re et al., 2012). In contrast, little attention has been paid to the degree to which therapists' overall ratings of their patients vary or cluster in comparison to the clustering of patient ratings of the same therapist. Studies conducted by Hatcher, Barends, Hansell & Gutfreund (1995) and Marcus, Kashy & Baldwin (2009) suggest that the clustering of patient ratings is much lower (0-6% of variance) compared to clustering of therapist ratings (20-30%).

Against this background, we explored in what way clinicians' ratings are affected if they rate their therapeutic relationship with more than one patient. Does this lead to a cluster effect, reducing the variance of ratings by a clinician, or does it increase the variance of ratings, because clinicians might want to emphasise differences between their patients? These questions can only be explored in data sets with patient and clinician ratings of similar constructs, such as the therapeutic relationship.

The aims of our research are to:

1. Compare the degree of clustering present in the rating score given by patients to that of clinicians.
2. Assess whether clinicians rate their therapeutic relationship with more variability than patients do, using a measure completed by both parties;

We have chosen two data sets in which to explore our questions, an observational study of communication between patients and their psychiatrists and a cluster randomised trial of regular feedback between patients and their key workers. Exploring our questions in two independent samples with a similar population and use

of the same rating scale gives the potential for a stronger finding should the results agree.

## **2. Methods**

We explored our research questions by conducting secondary analyses in two independent samples from longitudinal studies in which the Helping Alliance Scale (HAS) had been completed by both clinicians and patients. One was an observational study, the other one a randomised controlled trial.

### **2.1 Study description**

In the observational study (McCabe et al. 2013), various aspects of the patient-clinician communication were captured and analysed. The DIALOG study (Priebe et al. 2002, 2007) was a cluster randomised trial in six European countries aimed at improving patient quality of life. Details of the studies and their findings have been published previously (McCabe et al. 2013; Priebe et al. 2002; 2007). In both studies the therapeutic relationships were assessed by clinicians and patients on the HAS. Patient ratings were obtained in interviews with researchers who were not involved in the patients' care. All ratings, performed at baseline, were repeated after a 6 month follow up in the observational study and 12 month follow-up in the trial. In both of these studies, all patients had a diagnosis of schizophrenia or schizoaffective disorder. For the current study, we use clinician to mean both psychiatrists and key workers (i.e. a professional mental health worker who liaises with the patient).

### **2.2 Measures**

The HAS (Priebe & Gruyters, 1993) was devised to capture the essential elements of the relationship between patients and their clinicians in a community care setting. The five questions of the patient instrument, on an 11 point scale from 0 to 10, with 10 indicating the most positive response, are (1) Do you feel understood by your case manager? (2) Do you feel criticised by your case manager? (3) How much is your case

manager committed to and actively involved in the treatment? (4) Is the treatment you are currently receiving right for you? (5) Do you trust your case manager? The sum of these divided by 5 is calculated to give the HAS mean score. Clinicians use a different version of the HAS which has 11 point scales for the following questions: (1) Can you get along with this patient? (2) Can you understand this patient and his/her problems (3) Do you look forward to meeting this patient? (4) Do you feel actively involved in his/her care? (5) Do you feel you can help this patient? The mean score is given by the sum of these divided by 5. The HAS has good psychometric properties and has been widely used in research (Catty et al., 2010; Reininghaus et al., 2013). Whilst individual scale items differ between patient and clinician versions, the scale has demonstrated good discriminant validity, suggesting that different, but related concepts of the therapeutic relationship are assessed (Reininghaus et al., 2013).

### 2.3 Statistical approach

Descriptive statistics of the HAS from the two datasets are presented. Our main interest is in the spread of HAS scores within clinician, described using the within clinician variance. We also report the intraclass correlation coefficient (ICC) which estimates the correlation in scores between any pair of randomly selected patients under the care of a clinician or, conversely, the percentage of total variability in HAS due to variation between clinicians (Kerry & Bland, 1998).

$ICC = \frac{s_b^2}{s_b^2 + s_w^2}$ , where  $s_b^2$  is the variance between clusters (clinicians) and  $s_w^2$  is the variance within clusters (clinicians). We define a patient ICC i.e. of the therapeutic relationship rated by patients, and a clinician ICC i.e. of the therapeutic relationship rated by the clinicians.

The ICCs were estimated using linear mixed models with a random effect for clinician. We used both baseline and follow up data to explore whether findings at follow up reflected those at baseline. Analyses were performed in Stata version 12 (StataCorp., 2011). As the study data were drawn from different populations and study designs,

we did not examine statistics regarding differences in baseline characteristics, HAS ratings and ICCs.

### **3. Results**

In the observational study, 126 patients and their 28 clinicians both rated their therapeutic relationship using the HAS; 27 clinicians rated their therapeutic relationship at baseline with 123 patients, and 126 patients their therapeutic relationship with 28 clinicians. In the cluster randomised trial, data were collected on 134 clinicians of 507 patients; 132 clinicians rated their therapeutic relationship at baseline with 489 patients and 504 patients their therapeutic relationship with 134 clinicians. The majority of clinicians (82%) were male in the observational study whereas the majority were female (63%) in the trial. The average number of patients per clinician was greater in the observational study, and more variable, than in the trial. The mean age of patients completing education and percentage male were very similar between the two studies (table 1).

#### **3.1 Observational study**

The mean number of patients per clinician was 4.5 (s.d. = 4.4). One hundred and three patients were rated by their clinician at both baseline and follow-up. Conversely, 20 clinicians were rated by their patients at both baseline and follow-up.

Patients tended to score the HAS more favourably (mean = 8.2, s.d. = 2.0), than clinicians (mean = 7.4, s.d. = 1.7), i.e. on average 11% higher. The spread of HAS ratings was greater in patients than in clinicians. Similar observations were made at follow up (table 2).

The ICC of clinicians rating the HAS was 0.65 (95% CI: 0.48 to 0.79) indicating that there was significant clustering present in their ratings. The ICC of patients rating the HAS, on the other hand, was approximately 0. This was due to the observation that the

between clinician variance of the HAS score for patient ratings was approximately 0. Findings at follow up were very similar (table 2).

The within clinician variance was considerably larger for patient ratings (3.86, s.e. = 0.49) compared to clinician ratings (0.88, s.e. = 0.13) indicating that patients rating the same clinician did so with more variability than the clinician rating the same set of patients. Again findings at follow up were very similar (table 3).

### 3.2 The DIALOG trial

The mean number of patients per clinician was 3.7 (s.d. = 2.7). Four hundred and thirty one patients were rated by their clinician at both baseline and follow-up. Conversely, 88 clinicians were rated by their patients at both baseline and follow-up.

Table 2 shows results in the DIALOG trial were similar to those seen in the observational study. Here, patients tended to score the HAS more favourably (mean = 8.0, s.d. = 1.7), than clinicians (mean = 7.5, s.d. = 1.3) i.e. on average 7% higher. The spread of HAS ratings was greater in patients than in clinicians at both baseline and follow-up.

At baseline, the ICC of patients rating the HAS was 0.11 (95% CI: 0.05 to 0.22) and of clinicians rating the HAS was 0.31 (95% CI: 0.22 to 0.42). Thus greater clustering was seen in the ratings by clinicians compared to ratings by patients and there was overlap in the confidence intervals only near their limits (table 2).

Taken together, the ICCs (0.65 and 0.31) suggest that over two thirds of the total variability in clinician-rated HAS scores was due to differences between clinicians, whilst this was one third for the trial. The difference seen in ICCs is largely accounted for by the substantial difference in within clinician variation, comparing clinician ratings to patient ratings. The within clinician variance for patient ratings (2.53, s.e. = 0.18) was more than two times greater than for clinician ratings (1.07, s.e. = 0.08). Again this shows that several patients rating the same clinician did so with more

variability than a clinician rating multiple patients. Findings at follow up were very similar (table 3).

In both studies, the within-clinician variability of HAS,  $s_w^2$ , rated by the clinicians was notably smaller than that rated by the patients. The larger  $s_w^2$ , is as a proportion of the total variability, the smaller the ICC will tend to be. The small ICC values observed for patient-rated HAS compared to clinician-rated HAS are therefore, a direct consequence of the large within-clinician variability of the patient-rated HAS scores.

#### **4. Discussion**

Patients, whilst tending to rate the therapeutic relationship notably more highly than their clinicians, also rated it with greater variability. Clinicians rated the relationship with several of their patients with lower variability than their patients did, and showed a cluster effect. These findings were replicated in two independent samples drawn on different designs, one experimental and the other observational, and were found in each sample at each of two points of time which were at least six months apart. Thus, the analyses provided a clear and consistent answer to the research question: patients tended to rate the relationship more highly than their clinicians but with greater variability. We had expected that patients would tend to rate the therapeutic alliance with their clinician highly, which was the case, but with less variability given the ceiling effect of high ratings. We had anticipated that clinicians would draw distinctions between their relationships with different patients and thus use a greater portion of the scale. This was not the case: whilst clinicians tended to rate their relationships with patients slightly lower, they rated with less variability, suggesting an avoidance of the extreme lower end of the scale.

The amount of clustering present in the rating scores by clinicians was considerably larger than in the rating scores by patients. The ICCs for patient ratings were close to zero in the observational study and just over 0.1 in the trial. This contrasts with the ICC of 0.7 found for clinician ratings in the observational study and just over 0.3 in the

trial. These differences in ICC values are seen mainly due to the higher within-clinician variance when comparing patient ratings to clinician ratings.

Given the difference in study designs, it would not have been appropriate to have adjusted for other factors, and thus the ICCs may have been over-estimated. Whilst the HAS demonstrates good discriminant validity (Reininghaus et al., 2013), the difference in question items to patients and therapists may have also contributed to the difference in means and thus reflect difference in content rather than overall rating of the relationship. It is also unclear to what extent the researchers conducting interviews with patients may have influenced ratings, and this was not examined in this study.

#### 4.1 Comparison with existing literature

Imel and colleagues drew upon two large healthcare maintenance organisation samples and used multilevel modelling to estimate intra therapist correlations in alliance, finding values of 0.02 and 0.07 (Imel et al. 2013). They gave the example of patient ratings of therapists to measure the quality of therapist performance. They also cited examples of counselling centre samples with small ICCs (0-0.06) (Hatcher et al., 1995; Marcus et al., 2009; Owen, Quirk, Hilsenroth & Rodolfa, 2011; Owen, Tao, Leach & Rodolfa, 2011; Owen, Tao & Rodolfa, 2010), secondary analyses of RCTs with larger ICCs (0.10) (Crits-Christoph et al., 2009; Zuroff et al., 2010) and three studies with ICCs of around 0.20 including one Norwegian study set in outpatient clinics (Nissen-Lie, Monsen & Ronnestad, 2010). The ranges of these ICCs are generally smaller than those we observed as they were estimated from models adjusted for various factors as opposed to our estimates which we did not adjust on account of the differences in study design and variables collected.

In their study of 101 patients, Junghan and colleagues found that the ICC in staff-rated therapeutic alliance was 0.67 compared to that in patient-rated therapeutic alliance of 0.49. (Junghan, Leese, Priebe & Slade, 2007) Our findings were similar in that the ICCs of staff ratings were greater than those of patient ratings, and were of

comparable magnitude. This study used the same measure of alliance (HAS) and with a similar sample.

Underwood et al. conducted a large cluster randomised trial of exercise for depression in residential and nursing care homes in England (Underwood et al. 2013). Health related quality of life was measured by EQ-5D, both self-rated, by residents who had the capacity to do so, and by their carer. It was noted that the ICC of self-rated EQ-5D at 12 months post randomisation was considerably smaller than that rated by carers (0.02 compared to 0.17). This contrast in size also accords with what we found, although the therapeutic relationship we report is a construct of a shared relationship between clinician and patient rather than being focused on the patient alone.

In a meta-analysis, Del Re and colleagues examined therapist effects in the alliance-outcome correlation and found the ratio of patients to therapists to be a significant moderator when controlling for a range of confounders, including the research design (Del Re et al., 2012). The meta-analysis focused upon patient ratings of the alliance only and thus does not provide therapist ratings as a comparator. Their meta-analysis corroborates findings from previous studies (Baldwin et al., 2007; Dinger et al., 2008; Zuroff et al., 2010) leading them to conclude that therapist variability and thus therapist capacity to form a strong alliance plays the greatest role in contributing to patient outcomes. The above studies included patients with depression and reported relatively small ICCs (0.03-0.098). Our samples had a diagnosis of psychosis, but this difference in diagnoses may not be relevant in this context, as Del Re et al. (2012) did not find diagnosis to be a significant confounder in their meta-analysis.

Whilst clinicians tended to rate the alliance slightly lower than patients, they did so with less variability. This suggests an avoidance of the lower extreme ends of the scale, or a tendency to avoid rating the relationship as poor. This points to bias in terms of potential 'halo' effect (tendency to rate relationships more positively), social desirability (for example, wishing to rate a generally good relationship across all patients) or consistency effects.

The ratings made by one clinician on several patients are unlikely to be independent, violating an assumption of simple statistical methods, and this has long been recognised in the literature on, for example, clinical trials (Roberts, 1999) and comparative studies of psychotherapies (Crits-Christoph & Mintz, 1991; Elvins & Green, 2008). This has implications both for the planning and interpretation of studies. At the planning stage, considering the cluster effect identified here is likely to require a larger sample size to achieve the same statistical power (Walwyn & Roberts, 2010) and avoid an increased type I error rate (Lee & Thompson, 2005). Elvins & Green (2008) suggest mitigation of client and therapist inter-correlation via an observer measurement of the alliance. However, the findings from this study raise the question as to whether any rating by a single observer (such as a research assistant) may also be prone to such clustering effects and thus require accounting for in the planning and analysis of studies.

#### 4.2 Conclusion

In two independent samples, patients rated the therapeutic relationship more favourably than their clinicians. However, contrary to expectations, on average they used more of the rating scale than the clinicians. In the observational study, over two thirds of the total variability in clinician-rated HAS scores was due to differences between clinicians, and in the trial it was one third. These findings help us quantify and understand the different perspectives on the therapeutic relationship that patients and their clinicians may have. In both cases clinicians and those using the data should be aware that there is a cluster effect of clinician ratings which may impact on the scores, possibly by overemphasising differences between clinicians that cannot be discerned if only considering patient ratings. These findings may well apply to other measures where the patient and clinician assess the same or a similar construct.

The findings of this paper point towards a new aspect in the analysis of outcome data. When outcome data for several patients are obtained by the same clinician, a cluster effect may have to be considered for the ratings, and this is likely to influence sample

size calculations when planning studies or the statistical power when analysing them or both.

**Conflict of Interest Statement**

The authors have no conflict of interest to declare

## References

Ardito, R. B., & Rabellino, D. (2011). Therapeutic alliance and outcome of psychotherapy: Historical excursus, measurements, and prospects for research. *Frontiers in Psychology, 2*, 270. doi: 10.3389/fpsyg.2011.00270

Baldwin, S. A., Wampold, B. E., & Imel, Z. E. (2007). Untangling the alliance–outcome correlation: Exploring the relative importance of therapist and patient variability in the alliance. *Journal of Consulting and Clinical Psychology, 75*, 842–852. [doi:10.1037/0022-006X.75.6.842](https://doi.org/10.1037/0022-006X.75.6.842)

Catty, J., Koletsis, M., White, S., Becker, T., Fioritti, A., Kalkan, R., Lauber, C. et al. (2010). Therapeutic relationships: their specificity in predicting outcomes for people with psychosis using clinical and vocational services. *Social Psychiatry and Psychiatric Epidemiology, 45*, 1187-93. doi:10.1007/s00127-009-0163-9

Crits-Christoph, P., Gallop, R., Temes, C. M., Woody, G., Ball, S. A., Martino, S., & Carroll, K. M. (2009). The alliance in motivational enhancement therapy and counselling as usual for substance use problems. *Journal of Consulting and Clinical Psychology, 77*, 1125-1135. doi:10.1037/a0017045

Crits-Christoph, P., & Mintz, J. (1991). Implications of therapist effects for the design and analysis of comparative studies of psychotherapies. *Journal of Consulting and Clinical Psychology, 59*, 20-26. <http://dx.doi.org/10.1037/0022-006X.59.1.20>

Del Re, A. C., Flückiger, C., Horvath, A. O., Symonds, D., & Wampold, B. E. (2012). Therapist effects in the therapeutic alliance–outcome relationship: A restricted–maximum likelihood meta-analysis. *Clinical Psychology Review, 32*, 642-649. doi:10.1016/j.cpr.2012.07.002

Dinger, U., Strack, M., Leichsenring, F., Wilmers, F., & Schauenburg, H. (2008). Therapist effects on outcome and alliance in inpatient psychotherapy. *Journal of Clinical Psychology, 64*, 344–354. doi:10.1002/jclp. 20443

Elvins, R., & Green, J. (2008). The conceptualization and measurement of therapeutic alliance: An empirical review. *Clinical Psychology Review*, 28, 1167-1187. doi:10.1016/j.cpr.2008.04.002

Hatcher, R. L., Barends, A., Hansell, J., & Gutfreund, M. J. (1995). Patients' and therapists' shared and unique views of the therapeutic alliance: An investigation using confirmatory factor analysis in a nested design. *Journal of Consulting and Clinical Psychology*, 63, 636-643. doi:10.1037/0022-006X.63.4.636

Horvath, J. (2000). The therapeutic relationship: From transference to alliance. *Journal of Clinical Psychology*, 56, 163-173. doi:10.1002/(SICI)1097-4679(200002)56:2<163::AID-JCLP3>3.0.CO;2-D

Imel, Z. E., Hubbard, R. A., Rutter, C. M., & Simon, G. (2013). Patient-rated alliance as a measure of therapist performance in two clinical settings. *Journal of Consulting and Clinical Psychology*, 81, 154-165. doi: 10.1037/a0030903

Junghan, U. M., Leese, M., Priebe, S., & Slade, M. (2007). Staff and patient perspectives on unmet need and therapeutic alliance in community mental health services. *British Journal of Psychiatry*, 191, 543-547. doi: 10.1192/bjp.bp.107.037978

Kahan, B., & Morris, T. (2013). Assessing potential sources of clustering in individually randomised trials. *BMC Medical Research Methodology*, 13, 58. doi: 10.1186/1471-2288-13-58

Kerry, S. M., & Bland, J. M. (1998). The intraclass correlation coefficient in cluster randomisation. *British Medical Journal*, 316, 1455-1460. doi:10.1136/bmj.316.7142.1455

Lee, K. J., & Thompson, S. G. (2005). The use of random effects models to allow for clustering in individually randomized trials. *Clinical Trials*, 2, 163-173. doi: 10.1191/1740774505cn082oa

Marcus, D. K., Kashy, D. A., & Baldwin, S. A. (2009). Studying psychotherapy using the one-with-many design: The therapeutic alliance as an exemplar. *Journal of Counselling Psychology*, 56, 537-548. doi: 10.1037/a0017291

McCabe, R., Lavelle, M., Dodwell, D., Healey, P., Laugharne, R., Priebe, S., Snell, A., & Bremner, S. (2013). Shared understanding in psychiatrist-patient communication: Association with treatment adherence in schizophrenia. *Patient Education and Counseling*, 93, 73-79. [doi:10.1016/j.pec.2013.05.015](https://doi.org/10.1016/j.pec.2013.05.015)

Nissen-Lie, H. A., Monsen, J. T., & Ronnestad, M. H. (2010). Therapist predictors of early patient-rated working alliance: A multilevel approach. *Psychotherapy Research*, 20, 627-646. doi:10.1080/10503307.2010.497633

Owen, J. J., Quirk, K., Hilsenroth, M. J., & Rodolfa, E. (2011). Working through: In-session processes that promote between-session thoughts and activities. *Journal of Counseling Psychology*, 59, 161-167. doi:10.1037/a0023616

Owen, J. J., Tao, K., Leach, M. M., & Rodolfa, E. (2011). Clients' perceptions of their psychotherapists' multicultural orientation. *Psychotherapy*, 48, 274-282. doi:10.1037/a0022065

Owen, J. J., Tao, K., & Rodolfa, E. (2010). Microaggressions and women in short-term psychotherapy: Initial evidence. *The Counseling Psychologist*, 38, 923-946. doi:10.1177/0011000010376093

Phillips, L. A., Leventhal, E. A., & Leventhal, H. (2011). Factors associated with the accuracy of physicians' predictions of patient adherence. *Patient Education and Counseling*, 85, 461-467. doi:10.1016/j.pec.2011.03.012

Podsakoff, P. M., MacKenzie, S. B., & Lee, J-Y. (2003). Common method biases in behavioural research: A critical review of the literature and recommended remedies.

Journal of Applied Psychology, 88, 879-903. doi: 10.1037/0021-9010.88.5.879

Priebe, S., & Gruyters, T. (1993). The role of the helping alliance in psychiatric community care. A prospective study. *Journal of Nervous and Mental Disease*, 181, 552-557. doi: 10.1097/00005053-199309000-00004

Priebe, S., McCabe, R., Bullenkamp, J., Hansson, L., Lauber, C., Martinez-Leal, R., . . . Wright, D. (2007). Structured patient-clinician communication and 1-year outcome in community mental healthcare. *British Journal of Psychiatry*, 191, 420-426. doi: 10.1192/bjp.bp.107.036939

Priebe, S., McCabe, R., Bullenkamp, J., Hansson, L., Rossler, W., Torres-Gonzales, F., & Wiersma, D. (2002). The impact of routine outcome measurement on treatment processes in community mental health care: approach and methods of the MECCA study. *Epidemiology and Psychiatric Sciences*, 11, 198-205. doi:10.1017/S1121189X00005728

Reininghaus, U., McCabe, R., Slade, M., Burns, T., Croudace, T., & Priebe, S. (2013). The validity of patient- and clinician-rated measures of needs and the therapeutic relationship in psychosis: A pooled analysis. *Psychiatry Research*, 209, 711-720. doi:10.1016/j.psychres.2013.01.013

Roberts, C. (1999). The implications of variation in outcome between health professionals for the design and analysis of randomized controlled trials. *Statistics in Medicine*, 18, 2605-2615. doi:10.1002/(SICI)1097-0258(19991015)18:19<2605::AID-SIM237>3.0.CO;2-N

StataCorp. 2011. *Stata Statistical Software: Release 12*. College Station, TX: StataCorp LP.

Streiner, D. L., & Norman, G. R. (2008). *Health measurement scales: A practical guide to their development and use* (4<sup>th</sup> ed.). Oxford: Oxford University Press.

Thompson, L., Howes, C., & McCabe, R. (2015). Effect of questions used by psychiatrists on therapeutic alliance and adherence. *British Journal of Psychiatry*, 209, 40-47. doi:10.1192/bjp.bp.114.151910

Underwood, M., Lamb, S. E., Eldridge, S., Sheehan, B., Slowther, A. M., Spencer, A., . . . Taylor, S. (2013). Exercise for depression in elderly residents of care homes: a cluster-randomised controlled trial. *Lancet*, 382, 41-59. doi: [10.1016/S0140-6736\(13\)60649-2](https://doi.org/10.1016/S0140-6736(13)60649-2)

Walwyn, R., & Roberts, C. (2010). Therapist variation within randomised trials of psychotherapy: implications for precision, internal and external validity. *Statistical Methods in Medical Research*, 19, 291-315. doi:10.1177/0962280209105017

Zuroff, D. C., Kelly, A. C., Leybman, M. J., Blatt, S. J., & Wampold, B. E. (2010). Between-therapist and within-therapist differences in the quality of the therapeutic relationship: Effects on maladjustment and self-critical perfectionism. *Journal of Clinical Psychology*, 66, 681-697. doi:10.1002/jclp.20683

Table 1 Baseline characteristics of the clinicians and patients in the observational study and DIALOG trial

		Observational study	DIALOG trial
Characteristic		Mean (s.d.) or n (%)	Mean (s.d.) or n (%)
Clinicians	Age (years)	n/a	43.8 (8.7)
	Male	23/28 (82%)	52/139 (37%)
	Patients per clinician	4.5 (4.4)	3.7 (2.7)
Patients	Age finished education (years)	18.1 (4.2)	19.1 (5.1)
	Male	81/126 (64%)	336/507 (66%)
	Schizophrenia/schizoaffective	126/126 (100%)	427/507 (84%)

s.d. standard deviation, n/a not available

Table 2 Mean (SD) of HAS scores and intraclass correlation coefficients in the observational study and DIALOG trial

		HAS in the observational study (103 ratings)		HAS in the DIALOG trial (431 ratings)	
		Mean (s.d.)	ICC (95% CI)	Mean (s.d.)	ICC (95% CI)
Patient ratings	Baseline	8.2 (1.97)	0 (.)	8.0 (1.69)	0.11 (0.05 to 0.22)
	Follow up	8.0 (1.94)	0 (.)	8.0 (1.75)	0.15 (0.08 to 0.26)
Clinician ratings	Baseline	7.4 (1.67)	0.65 (0.48 to 0.79)	7.5 (1.26)	0.31 (0.22 to 0.42)
	Follow up	7.2 (1.78)	0.71 (0.55 to 0.83)	7.5 (1.16)	0.31 (0.22 to 0.43)

s.d. standard deviation, ICC intraclass correlation coefficient, CI confidence interval, HAS Helping Alliance scale

Table 3 Estimates (SEs) of the between and within clinician variances of HAS scores in the observational study and DIALOG trial

		HAS in the observational study (103 ratings)		HAS in the DIALOG trial (431 ratings)	
		Between clinician variance (s.e.)	Within clinician variance (s.e.)	Between clinician variance (s.e.)	Within clinician variance (s.e.)
Patient ratings	Baseline	0 (.)	3.86 (0.49)	0.30 (0.13)	2.53 (0.18)
	Follow up	0 (.)	3.74 (0.51)	0.45 (0.15)	2.59 (0.20)
Clinician ratings	Baseline	1.64 (0.52)	0.88 (0.13)	0.48 (0.10)	1.07 (0.08)
	Follow up	2.20 (0.68)	0.90 (0.13)	0.42 (0.09)	0.91 (0.07)

s.e. standard error