DOI:
10.1002/mpr.1480

Document Version
Peer reviewed version

Cite as: Pesola F, Williams J, Bird V, Freidl M, Le Boutillier C, Leamy M, Macpherson R, Slade M (2015) *Development and evaluation of an Individualised Outcome Measure (IOM) for randomised controlled trials in mental health*, International Journal of Methods in Psychiatric Research, **24**, 257-265.

Development and evaluation of an Individualised Outcome Measure (IOM) for randomised controlled trials in mental health

Francesca Pesola[1]

Julie Williams[1]

Victoria Bird[1]

Marion Freidl[12]

Clair Le Boutillier[1]

Mary Leamy[1]

Rob Macpherson[3]

Mike Slade[1]

[1] King's College London, Health Service and Population Research Department, Institute of Psychiatry, Psychology & Neuroscience, Denmark Hill, London SE5 8AF, UK

[2] Clinical Division of Social Psychiatry, Department of Psychiatry and Psychotherapy, Medical University of Vienna, Austria

[3] Leckhampton Lodge, Charlton Lane, Cheltenham GL53 9DZ

Short title: Individualised Outcome Measures for RCTs

Corresponding Author:

Mike Slade. Tel: 020 7848 5095, Email: mike.slade@kcl.ac.uk, King's College London, Institute of Psychiatry, Psychology & Neuroscience, Section for Recovery (Box P029), Health Service and Population Research Department, De Crespigny Park, London, SE5 8AF, UK.

Target journal – International Journal of Methods in Psychiatric Research

Abstract = 198 (max. 200)

Words = 4,044 (max. 5000)

Tables = 5

Abstract

Predefined, researcher-selected outcomes are routinely used as the clinical end-point in randomised controlled trials (RCTs); however, individualised approaches may be an effective way to assess outcome in mental health research. The present study describes the development and evaluation of the Individualised Outcome Measure (IOM), which is a patient-specific outcome measure to be used for RCTs of complex interventions. IOM was developed using a narrative review, expert consultation and piloting with mental health service users (n=20). The final version of IOM comprises two components: Goal Attainment (GA) and Personalised Primary Outcome (PPO). For GA, patients identify one relevant goal at baseline and rate its attainment at follow-up. For PPO, patients choose an outcome domain related to their goal from a predefined list at baseline, and complete a standardised questionnaire assessing the chosen outcome domain at baseline and follow-up. A feasibility study indicated that IOM had adequate completion (89%) and acceptability (96%) rates in a clinical sample (n=84). IOM was then evaluated in an RCT (ISRCTN02507940). GA and PPO components were associated with each other and with the trial primary outcome. The use of the PPO component of IOM as the primary outcome could be considered in future RCTs.

Keywords: randomised controlled trial, mental health, personalised measure, individualised outcome

Introduction

Randomised controlled trials (RCTs) evaluate the effectiveness of an intervention in relation to a pre-specified primary outcome (Schulz et al., 2010). Predefining a standardised measure as a primary outcome has several advantages, including informing the power calculation for the sample size, reducing post-hoc 'data fishing', and supporting meta-analytic aggregation of findings across studies. However, the use of a single predefined primary outcome to evaluate complex interventions raises four issues.

First, a range of outcome domains exist (Slade, 2002), and as complex interventions may impact across this range (Craig et al., 2008), a single predefined outcome measure may not capture all changes. Second, some measures may become widely used primarily because they have been used in other RCTs, rather than because of any inherent superiority (Everitt and Wessely, 2004). Third, pre-specifying the primary outcome for an RCT may not be clinically informative, as recognised in the developing field of clinimetrics (Feinstein, 1983). Finally, patients may vary in the importance they attach to different types of outcomes. For example, the aim of treatment for some people may be symptomatic improvement, whereas others have social goals (*e.g.* employment) or personal goals (*e.g.* to find meaning in their life) (Slade, 2009). As a result, a predefined outcome measure may not be relevant to the person receiving the intervention.

A number of patient-centred measures and approaches have been developed to address these issues, such as Goal Attainment Scaling (GAS) and Target Complaint Approach (Donnelly and Carswell, 2002). The GAS process was originally developed for mental health programmes (Kiresuk and Sherman, 1968) and it involves the patient, together with the clinician, prospectively identifying a number of relevant goals, and then rating their progress towards each goal along a five-point scale at follow-up. Although each respondent has different goals, the overall GAS score is standardised so that data can be aggregated and compared across individuals (McDougall and King, 2007; Turner-Stokes, 2009). GAS has been effectively used to identify and evaluate outcomes in RCTs in rehabilitation (Graven et al., 2011; Turner-Stokes et al., 2010) and studies with older adults (Rockwood et al., 2007; Rockwood et al., 2003). Systematic reviews conclude that GAS shows high sensitivity to change, reliability and validity when used in chronic healthcare and physical and

neurological rehabilitation (Hurn et al., 2006; Stevens et al., 2013). A systematic review of research evaluating the implementation of GAS in occupational health programmes found that goal setting could be implemented in mental health settings and that goal attainment was associated with positive change in clinical outcomes (Scott and Haggarty, 1984). GAS was also found to be an effective clinical tool among patients in psychiatric day care, with patients with schizophrenia having the highest success rate with 75% achieving a good outcome (Falloon and Talbot, 1982).

Despite this evidence, fully individualised goal setting is not implemented in randomised controlled trials in the mental health field. This may be due to operational limitations associated with GAS (Ertzgaard et al., 2011; King et al., 1999). In particular, goal scaling relies on a complex formula, making calculations complex and interpretation complicated (Stevens et al., 2013). Moreover, GAS is based on scaling at baseline and follow-up which may not fully capture the actual level of goal attainment. Finally, researchers would most likely be involved in goal selection and scaling in RCTs, and they may not be able to assess whether the goal is realistic at baseline or rate attainment at follow-up.

The aim of this study was to develop and evaluate a new trans-diagnostic and patient-valued approach to assessing change in RCTs of complex interventions in mental health, by using individualised and positively-valued outcome domains. In line with recovery research (Leamy et al., 2011), the focus was on subjective experiental outcomes rather than clinical outcomes. The objectives were to develop, pilot, and assess the feasibility of the Individualised Outcome Measure (IOM), and then to evaluate IOM in an RCT by assessing whether IOM components 1) capture change over time, 2) co-vary, and 3) are associated with the predefined primary outcome.

Method

Participants

Participants were recruited from community-based mental health teams in the South London and Maudsley NHS Foundation Trust (pilot, feasibility, RCT) and 2Gether NHS Foundation Trust (RCT). Inclusion criteria were being aged 18-65 years, being able to speak

and understand English, and being sufficiently well to participate in the opinion of the clinician providing community support. In the RCT, patients were participating in the REFOCUS Trial (ISRCTN02507940) which had additional inclusion criteria. These were having a primary clinical diagnosis of psychosis, having no immediate plans for discharge or transfer, not currently receiving in-patient care, not being in prison, not participating in another substantial study, and being in regular contact with at least one worker in the team. Ethical approval was obtained (10/H0807/4).

Measures

The final version of the Individualised Outcome Measure (IOM Version 3) comprised two components to index individual outcomes: Goal Attainment (GA) and Personalised Primary Outcome (PPO). The GA component is a variation of GAS, requiring patients to identify one personally relevant goal which mattered to them, that they thought they could achieve and which services could help with, at baseline. Patients were asked to identify only one goal as evidence from mental health research showed respondents can focus on one goal at a time (Scott and Haggarty, 1984). Having one goal also simplifies assessment of change and aids interpretability of the findings. At follow-up, the patient rated how successful they were at reaching their goal on a 5-point Likert scale from 0 (I am further away from my goal) to 4 (I did even better than expected).

After identifying their goal, patients were invited to complete the PPO component, which is a novel approach. It requires patients to choose the outcome domain which most closely maps onto their selected goal from a predefined list. Each domain is associated with one standardised measure, which patients complete at baseline and again at follow-up.

In IOM Version 1, thirteen outcome domains and associated patient-rated measures were included in the PPO component. The **Activity and Participation Questionnaire** (APQ) is a 5-item measure of social inclusion ranging from 0 to 24 (Coombs et al., 2013). The **Short Form Health Survey** (SF-36) is a 36-item measure of functional health and well-being, consisting of 8 scaled scores each with a score range from 0 to 100 (Ware and Sherbourne, 1992). The **Short Health Survey Questionnaire** (SF-12) is a 12-item measure of functional health and well-being, consisting of 2 scaled scores with a score range from 0 to 100 (Jenkinson and

Layte, 1997). The **Empowerment Scale** (ES) is a 28-item measure ranging from 1 to 112 (Rogers et al., 1997). The **Herth Hope Index** (HHI) is a 12-item measure of hope ranging from 1 to 48 (Herth, 1992). The **Rosenberg Self-Esteem Scale** (RSES) is a 10-item measure ranging from 1 to 40 (Rosenberg, 1965), items were recoded so that high scores indicated higher self-esteem levels. The **Stigma Scale** (SS) is a 28-item measure of stigma of mental illness and ranges from 1 to 140, with high scores indicating low stigma (King et al., 2007). The **Meaning of Life Questionnaire** (MLQ) is a 10-item measure ranging from 1 to 70 (Steger et al., 2006). MLQ comprises two subscales: Search for and Presence of meaning of life, and only the Presence subscale was used. The **MOS Social Support Survey** (MOS) is a 21-item measure ranging from 1 to 85 (Sherbourne and Stewart, 1991). The **Community Integration Measure** (CIM) is a 10-item measure ranging from 1 to 50 (McColl et al., 2001). The **Warwick-Edinburgh Mental Well-Being Scale** (WEMWBS) is a 14-item measure of wellbeing ranging from 1 to 70 (Tennant et al., 2007). The **Service user Perception of Functioning Scale** (PPFS) is a 6-item measure with ratings for both community functioning and cognition, rated ranging from 1 to 30 (Horn et al., 2012). The **Manchester Short Assessment of Quality of Life** (MANSA) is a 12-item measure ranging from to 84 (Priebe et al., 1999).

Two additional measures were used. The **Mental Health Confidence Scale** (MHCS) is a 16-item measure of empowerment ranging from 1 to 96 with good psychometric properties (Carpinello et al., 2000; Castelein et al., 2008). The **Questionnaire about the Process of Recovery** (QPR) was the primary outcome of the REFOCUS trial (Neil et al., 2009; Slade et al., 2011). This is a 22-item patient-rated assessment of recovery, with each item rated on a five-point scale from 0 to 4. An overall score was extrapolated by summing 15 of the original 22 items (Law H et al., 2014). All scales were coded so that a high score indicated a positive outcome.

The two-stage development process

Development and piloting

The PPO was developed using a narrative review of outcome domains in mental health (Bjorkman and Svensson, 2005; Myers et al., 2005; National Institute for Mental Health in England, 2008; Rush Jr, 2008; Slade, 2009) with specific emphasis on domains valued by patients (Faulkner and Layzell, 2000; Perkins and Repper, 2003; Rosenheck et al., 2005). For

each domain, a related outcome measure was sought. This produced the initial version of IOM Version 1, which was consulted on with the REFOCUS advisory boards, comprising a Lived Experience Advisory Panel (LEAP) of service users and carers, topic-specific experts, a virtual advisory panel of service users, researchers and other stakeholders with an interest in black and minority ethnic mental health, and an International Advisory Board of international experts. The consultation focused on identifying outcome domain omissions, measures for outcome domains which did not have identified measures as well as to improve the content, order and format of the PPO, which consisted of 13 outcome domains and corresponding measures. Following this consultation, IOM Version 2 was finalised.

A pilot study was, subsequently, undertaken with a convenience sample of 20 patients, to assess their experience of completing IOM Version 2. Characteristics of pilot study participants are shown in Table 1.

Insert Table 1 here

The pilot was conducted to identify any necessary change which may improve IOM. Informed consent was obtained and patients were paid £10 for their participation. All patients were able to complete and provide information on the IOM. Following the pilot, three measures of the PPO (APQ, SF-36 and SF-12) were deleted as they were not chosen by any respondent. Additionally, the Empowerment Scale was replaced with the Mental Health Confidence Scale. No changes were necessary for the GA component. Table 1 shows the list of ten outcome domains and their associated measures which constituted IOM Version 2.

Insert Table 2 here

Feasibility

We then assessed the feasibility of IOM Version 3 using a convenience sample of 84 current mental health patients, whose characteristics are shown in Table 1. They completed IOM at baseline and at 3-month follow-up, and rated their experience of completing IOM. Participants were paid £10 for each round of assessment.

A total of 75 (89%) patients were able to identify a goal, a relevant outcome domain, and complete the associated PPO measure. The time to identify a goal ranged between 1 and 5 minutes (median = 1, iqr = 1). We found that 64 (85%) patients reported the goal was 'very' or 'extremely' important to them while only 3 (4%) were not happy about being asked. All ten PPO domains and their associated measures were selected and completed (Table 2, column 4). At follow-up, 55 (73%) patients completed both IOM components. There was no association between attrition and any of the sociodemographic variables. For the GA component, 8 (15%) rated the goal as Fully Achieved, 28 (51%) as Partly Achieved and 19 (35%) as Not Achieved. Finally, 36 (65%) individuals rated their original goal as still Very or Extremely important.

Evaluation in an RCT

We then evaluated IOM Version 3 as an outcome measure in the REFOCUS Trial. The trial evaluated a team-based intervention intended to impact on the content and process of care (Slade et al., 2011). The intervention was intended to lead to more collaborative staff-patient relationships, and greater staff-focus on patient values, strengths, and goal-striving. The intended benefits were more patient-centred and recovery-oriented care, and this emphasis on individualised care in this team-based and trans-diagnostic intervention meant that the use of a predefined and invariant primary outcome was problematic in capturing the outcome domain relevant to each patient. The QPR was the predefined primary outcome.

Participating teams (n=27) were randomly allocated to control (n=13, treatment as usual) or intervention (n=14, treatment as usual plus REFOCUS Intervention). A random sample of 15 patients was recruited from each participating team, and after giving informed consent, they completed a set of measures including QPR and IOM Version 3 at baseline and at one-year follow-up. Participants were paid £10 for each round of assessment.

Data analysis

The effectiveness of the intervention did not differ between arms (Slade et al., in press), so data from both trial arms were pooled to address our hypotheses. We explored the relationship between attrition (missing vs. present) and socioeconomic variables of gender

age, ethnicity, marital status, education and accommodation type for the feasibility study and the trial. Missing data across the outcome scales were pro-rated when less than 20% of items were left blank.

Prior to conducting our analysis, we recoded the IOM components to facilitate analysis. To avoid small cell counts, GA was recoded so that 'successful' and 'even better than expected' were coded as Fully Achieved, 'some progress' was coded as Partly Achieved, and 'no progress' and 'further away from the goal' were coded as Not Achieved. To combine scores of the PPO, we standardised each measure using z-scores based on population norms from previous research.

Objective 1 - change over time

In order to test if the difference was significantly different from 0, we conducted a regression analysis with no predictors on the PPO difference score between baseline and follow-up. Subsequently, we assessed whether change on the PPO was clinically meaningful by implementing the Reliable Change Index (RCI) approach following the Jacobson and Truax's guidelines (Jacobson and Truax, 1991):

(x2-x1)/sqrt(2*(s*sqrt(1-rxx))2)

where *rxx* is the Cohen's alpha (i.e., reliability) value and *s* is the scale standard deviation based on population values. RCI is a standardised measure of change and scores greater than the critical value (i.e., +/-1.96) correspond to reliable change (Zahra and Hedge, 2010). Using these cut-off points, we could group patients as having 'Improved', 'Not changed', or 'Declined'.

Objective 2 – co-variation

To assess whether the two IOM components are associated and tap onto the same construct, we regressed the PPO follow-up score on GA while adjusting for baseline standardised scores.

Objective 3 – relationship with primary outcome

We assessed the relationship of the two IOM component with the REFOCUS trial predefined primary outcome of QPR by regressing the QPR change score separately onto the PPO change scores and GA.

In all regression analyses, we accounted for clustering at the team level by conducting random effects regression analyses with maximum likelihood estimation using the 'xtmixed' command in Stata 11. Site and trial arm were entered as covariates in the model in order to reflect the study design. Bonferroni corrections were implemented to adjust for multiple pairwise comparisons when appropriate.

RESULTS

At baseline, 340 (86%) patient participants in the RCT completed both components of the IOM. Table 1 summarises their sociodemographic characteristics. Patients with complete IOM at baseline were more likely to be White British (vs. other; $chi^2_{(1)}$=8.5, p=.004), to live in private accommodation (vs. supported; $chi^2_{(1)}$=4.6, p=.03) and to have formal qualifications (vs. no qualification; $chi^2_{(1)}$=20.9, p<.001).

At one-year follow-up, 239 (of 340) patients completed both IOM components. Patients with complete information at follow-up were more likely to have formal qualifications (vs. no qualification; $chi^2_{(1)=}$6.7, p=.01) and to be younger ($t_{(374)}$=2.4, p=.02). All ten PPO domains and their associated measures were selected and completed (Table 1, column 5). The goal was rated as Fully Achieved by 63 (26%) patients, 113 (47%) rated it as Partly Achieved while 63 (26%) reported having made No Progress.

Objective 1: change over time

The regression analysis (average cluster size = 9, range: 1 to 13) indicated that the change score on the PPO between baseline and follow-up was not different from 0 (z=1.3, p=.18, n=239). The RCI approach indicated that 79% patients did not change over time (Table 3).

Insert Table 3 here

Objective 2: co-variation

Analyses were conducted on all 239 participants with complete information on the IOM. A regression analysis across all 27 clusters (average cluster size = 9, range 1 to 13) showed an association between GA and PPO ($\chi^2_{(2)}$=13.8, p=.001; n=239). Compared to patients who had made No Progress, those who had Partly (b=.32, 95%CI: .09 to .56, z=2.7, p=.007) or Fully Achieved their goal (b=.50, 95%CI: .23 to .77, z=3.7, p<.001) had higher PPO scores at follow-up, while adjusting for baseline scores and accounting for multiple testing.

Objective 3: relationship with primary outcome

Regression analysis on all 237 patients with complete information on IOM and QPR (average cluster size = 9, range 1 to 13) showed that the PPO and QPR change scores were positively associated (b=3.3, 95%CI: 2.3 to 4.4, z=6.2, p<.001). Similarly, regression analysis across all 27 clusters (average cluster size = 9, range 1 to 13) showed that GA was associated with follow-up QPR scores ($\chi^2_{(2)}$=12.4, p=.002). In particular, patients who had Partly (b=3.0, 95%CI: 0.8 to 5.2, z=2.6, p=.008) or Fully Achieved their goal (b=4.3, 95%CI: 1.9 to 6.8, z=3.4, p=.001) reported greater recovery scores than those who had Not Achieved their goals, even after adjusting for multiple testing.

DISCUSSION

Following growing consensus that individualised approaches may be an effective way to assess outcome in mental health research (Slade, 2009), a new measure of individualised outcome for use as a clinical end-point in RCTs was developed, by literature review, expert consultation, piloting, assessment of feasibility and then evaluation in the context of an RCT. The IOM has two components, both of which had adequate acceptability and completion rates. Evaluation in an RCT showed that IOM could be used to differentiate between patients who changed versus those who did not change, that the two components (GA and PPO) co-vary, and that both were associated with change in the predefined primary outcome for the trial.

For GA, the majority of respondents (66% feasibility study, 72% RCT) reported having made at least some progress on their goal at follow-up. These figures are in line with those observed in previous studies in mental health and rehabilitation (Hurn et al., 2006; Scott and Haggarty, 1984; Stevens et al., 2013). For example, 63% of patients with schizophrenia were

found to have made good progress on their goal at 6-month follow-up (Falloon and Talbot, 1982). These results suggest that the GA component, which is a variation of the GAS, is as effective as the original version at capturing achievement.

The PPO component did not capture any statistical or clinically meaningful change across time (objective 1). Nonetheless, no change was observed on the REFOCUS trial primary outcome (Slade et al., in press), which suggests that change did not occur in the 12-month period. Hence, implementation and replication in a future trial is needed. PPO scores at follow-up were associated with Goal Attainment (objective 2), which indicates the two IOM components tap onto overlapping domains as originally intended. The PPO change scores were positively associated with change scores on QPR (objective 3). Thus, it appears to capture change levels in similar fashion to that exhibited by a standardised questionnaire.

Finally, patients who had Fully or Partly achieved their goal were more likely to report improved QPR scores than those who had not achieved their goal (objective 3). These results are in line with evidence that GAS scores were positively associated with standardised measures of progress in rehabilitation (de Beurs et al., 1993; Hurn et al., 2006).

One limitation is the fact that we cannot ascertain whether goal attainment leads to a greater sense of recovery or whether increased recovery leads to goal attainment as GA was assessed at follow-up only. Previous research indicates that goal attainment predicts progression in recovery (Clarke et al., 2009), and future studies implementing a mixed-method design may help us to further understand this association. Patients were assumed to be failing at baseline but this could not be corroborated which represents a limitation. Future studies may incorporate baseline assessment to index change on this component, although this would require more complex calculations.

Combining the different measures of the PPO list may be criticised for potentially leading to biased results; however, we used a norm-based standardisation process which has been found to overcome these issues (Van Cleave et al., 2011). Finally, even though variations of the original RCI exists which take into account the regression to the mean phenomenon (Wise, 2004), research showed that the original RCI approach represents an optimal

measure of change (Marsden et al., 2011; McGlinchey et al., 2002). The main strength of the study was the large and clinically representative sample recruited for an RCT in mental health, which renders our findings generalisable.

In summary, analysis of feasibility and evaluation of the IOM indicated that this approach can be implemented as an individualised outcome in RCTs for complex interventions in mental health. As the PPO provides an ecologically valid, standardised and more easily comparable approach to outcome measurement, we recommend the PPO component be retained, with the addition of the goal-identification element of GA as selecting a goal facilitates identifying a relevant outcome domain. Future research should further assess the psychometric properties (e.g., internal consistency, construct validity and test-retest reliability) of the PPO and establish whether it could be implemented in different types of intervention in mental health research.

Acknowledgement

This article presents independent research funded by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research scheme (RP-PG-0707-10040) and supported by the NIHR Mental Health Research Network in the context of the NIHR Biomedical Research Centre for Mental Health at South London and Maudsley NHS Foundation Trust and [Institute of Psychiatry, Psychology & Neuroscience] King's College London. The project will be published in full in the NIHR PGfAR journal. The views expressed in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, the PGfAR programme or the Department of Health. Further information available at researchintorecovery.com/refocus.

Declaration of interest statement.

No conflict of interest.

REFERENCES

Bjorkman T, Svensson B. (2005). Quality of life in people with severe mental illness. Reliability and validity of the Manchester Short Assessment of Quality of Life (MANSA). *Nordic journal of psychiatry* **59**(4): 5.

Carpinello S, Knight E, Markowitz F, Pease E. (2000). The development of the mental health confidence scale: A measure of self-efficacy in individuals diagnosed with mental disorders. *Psychiatric Rehabilitation Journal* **23**: 236 - 243.

Castelein S, van der Gaag M, Bruggeman R, van Busschbach JT, Wiersma D. (2008). Measuring Empowerment Among People With Psychotic Disorders: A Comparison of Three Instruments. *Psychiat Serv* **59**(11): 1338-1342, DOI: DOI 10.1176/appi.ps.59.11.1338.

Clarke S, Oades G, Crowe T (2009). The Role of Symptom Distress and Goal Attainment in Assisting the Psychological Recovery in Consumers with Enduring Mental Illness. *Journal of Mental Health***, 18**, 389-397.

Coombs T, Nicholas A, Pirkis J. (2013). A review of social inclusion measures. *Aust N Z J Psychiatry* **47**(10): 14.

Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. (2008). Developing and evaluating complex interventions: the new Medical Research Council guidance. *British Medical Journal* **337**.

de Beurs E, Lange A, Blonk RB, Koele P, van Balkom ALM, Van Dyck R. (1993). Goal attainment scaling: An idiosyncratic method to assess treatment effectiveness in agoraphobia. *J Psychopathol Behav Assess* **15**(4): 357-373, DOI: 10.1007/bf00965038.

Donnelly C, Carswell A (2002). Individualised outcome measures: A review of the literature. *Canadian Journal of Occupational Therapy*, **69:** 84-94.

Ertzgaard P, Ward AB, Wissel J, Borg J. (2011). Practical considerations for goal attainment scaling during rehabilitation following acquired brain injury. *Journal of rehabilitation medicine* **43**(1): 8-14, DOI: 10.2340/16501977-0664.

Everitt B, Wessely S. (2004). *Clinical trials in psychiatry*. Oxford: Oxford University Press.

Falloon IR, Talbot RE. (1982). Achieving the goals of day treatment. *The Journal of nervous and mental disease* **170**(5): 279-285.

Faulkner A, Layzell S. (2000). Strategies for Living. A report of user-led research into people's strategies for living with mental distress. London.

Feinstein AR. (1983). An additional basic science for clinical medicine: IV. The development of clinimetrics. *Annals of internal medicine* **99**(6): 843-848.

Graven C, Brock K, Hill K, Ames D, Cotton S, Joubert L. (2011). From rehabilitation to recovery: protocol for a randomised controlled trial evaluating a goal-based intervention to reduce depression and facilitate participation post-stroke. *BMC neurology* **11**: 73, DOI: 10.1186/1471-2377-11-73.

Herth K. (1992). Abbreviated instrument to measure hope: development and psychiatric evaluation. *Journal of Advanced Nursing* **17**: 1251-1259.

Horn KK, Jennings S, Richardson G, van Vliet D, Hefford C, Abbott JH. (2012). The Patient-Specific Functional Scale: Psychometrics, Clinimetrics, and Application as a Clinical Outcome Measure. *J Orthop Sport Phys* **42**(1): 30-42, DOI: DOI 10.2519/jospt.2012.3727.

Hurn J, Kneebone I, Cropley M. (2006). Goal setting as an outcome measure: A systematic review. *Clinical rehabilitation* **20**(9): 756-772, DOI: 10.1177/0269215506070793.

Jacobson NS, Truax P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of consulting and clinical psychology* **59**(1): 12-19.

Jenkinson C, Layte R. (1997). Development and testing of the UK SF-12. *Journal of health services research and policy* **2**(1): 5.
King G, McDougall J, Palisano R, Gritzan J, Tucker MA. (1999). Goal Attainment Scaling: Its Use in Evaluating Pediatric Therapy Program. *Physical & Occupational Therapy in Pediatrics* **19**: 31-52.
King M, Dinos S, Shaw J, Watson R, Stevens S, Passetti F, Weich S, Serfaty M. (2007). The Stigma Scale: development of a standardised measure of the stigma of mental illness. *The British journal of psychiatry : the journal of mental science* **190**: 248-254, DOI: 10.1192/bjp.bp.106.024638.
Kiresuk TJ, Sherman RE. (1968). Goal attainment scaling: A general method for evaluating comprehensive community mental health programs. *Community mental health journal* **4**(6): 443-453, DOI: 10.1007/bf01530764.
Law H, Neil S, Dunn G, AP M. (2014). Psychiatric properties of the questionnaire about the process of recovery (QPR) *Schizophr Res* **156**(2-3): 184-189.
Leamy M, Bird V, Le Boutillier C, Williams J, Slade M (2011) A conceptual framework for personal recovery in mental health: systematic review and narrative synthesis, *British Journal of Psychiatry*, **199**, 445-452.
Marsden J, Eastwood B, Wright C, Bradbury C, Knight J, Hammond P. (2011). How best to measure change in evaluations of treatment for substance use disorder. *Addiction (Abingdon, England)* **106**(2): 294-302, DOI: 10.1111/j.1360-0443.2010.03143.x.
McColl MA, Davies D, Carlson P, Johnston J, Minnes P. (2001). The community integration measure: development and preliminary validation. *Archives of physical medicine and rehabilitation* **82**(4): 6.
McDougall J, King G. (2007). Goal Attainment Scaling: Description, Utility, and Applications in Pediatric Therapy Services. (Archived by WebCite® at http://www.webcitation.org/6TXFJebT2). London, Ontario.
McGlinchey JB, Atkins DC, Jacobson NS. (2002). Clinical significance methods: Which one to use and how useful are they? *Behavior Therapy* **33**(4): 529-550, DOI: http://dx.doi.org/10.1016/S0005-7894(02)80015-6.
Myers F, McCollam A, Woodhouse A. (2005). Equal Minds. National Programme for Improving Mental Health and Wellbeing. Edinburgh: Scottish Government.
National Institute for Mental Health in England. (2008). *Outcomes compendium*. Birmingham: NIMHE.
Neil S, Kilbride M, Pitt L, nothard S, Welford M, Sellwood W, Morrison A. (2009). The questionnaire about the process of recovery (QPR): A measurement tool developed in collaboration with service users. *Psychosis* **1**: 145-155.
Perkins R, Repper J. (2003). *Social Inclusion and Recovery*. London BailliŠre Tindall
Priebe S, Huxley P, Knight S, Evans S. (1999). Application and results of the Manchester Short Assessment of Quality of Life (MANSA). *The International journal of social psychiatry* **45**(1): 7-12.
Rockwood K, Fay S, Jarrett P, Asp E. (2007). Effect of galantamine on verbal repetition in AD: a secondary analysis of the VISTA trial. *Neurology* **68**(14): 1116-1121, DOI: 10.1212/01.wnl.0000258661.61577.b7.
Rockwood K, Howlett S, Stadnyk K, Carver D, Powell C, Stolee P. (2003). Responsiveness of goal attainment scaling in a randomized controlled trial of comprehensive geriatric assessment. *Journal of clinical epidemiology* **56**(8): 736-743.

Rogers E, Chamberlin J, Ellison M, Crean T. (1997). A consumer-constructed scale to measure empowerment among users of mental health services. *Psychiat Serv* **48**(8): 6.
Rosenberg M. (1965). *Society and the Adolescent Self-Image*. Princeton, NJ: Princeton University Press.
Rosenheck R, Stroup S, Keefe R, McEvoy J, Swartz M, Perkins D. (2005). Measuring outcome priorities and preferences in people with schizophrenia. *British Journal of Psychiatry* **187**: 8.
Rush Jr AJ. (2008). *Handbook of Psychiatric Measures*. Washington: American Psychiatric Publishing.
Schulz KF, Altman DG, Moher D. (2010). CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *Obstetrics and gynecology* **115**(5): 1063-1070, DOI: 10.1097/AOG.0b013e3181d9d421.
Scott AH, Haggarty EJ. (1984). Structuring Goals via Goal Attainment Scaling in Occupational Therapy Groups in a Partial Hospitalization Setting. *Occupational Therapy in Mental Health* **4**(2): 39-58, DOI: 10.1300/J004v04n02_04.
Sherbourne C, Stewart A. (1991). The MOS social support survey. *Social Science and Medicine* **32**(6): 8.
Slade M, Bird V, Clarke E, Le Boutillier C, McCrone P, Macpherson R, Pesola F, Wallace G, Slade M. (2002). What outcomes to measure in routine mental health services, and how to assess them: a systematic review. *Aust N Z J Psychiatry* **36**(6): 743 - 753.
Slade M. (2009). *Personal recovery and mental illness. A guide for mental health professionals*. Cambridge: Cambridge University Press.
Slade M, Bird V, Le Boutillier C, Williams J, McCrone P, Leamy M. (2011). REFOCUS Trial: protocol for a cluster randomised controlled trial of a pro-recovery intervention within community based mental health teams. *BMC psychiatry* **11**: 185, DOI: 10.1186/1471-244x-11-185.
Steger MF, Frazier P, Oishi S, Kaler M. (2006). The meaning in life questionnaire: Assessing the presence of and search for meaning in life. *Journal of Counseling Psychology* **53**(1): 80-93, DOI: 10.1037/0022-0167.53.1.80.
Stevens A, Beurskens A, Koke A, van der Weijden T. (2013). The use of patient-specific measurement instruments in the process of goal-setting: a systematic review of available instruments and their feasibility. *Clinical rehabilitation* **27**(11): 1005-1019, DOI: 10.1177/0269215513490178.
Tennant R, Hiller L, Fishwick R, Platt S, Joseph S, Weich S, Parkinson J, Secker J, Stewart-Brown S. (2007). The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. *Health and quality of life outcomes* **5**: 63, DOI: 10.1186/1477-7525-5-63.
Turner-Stokes L. (2009). Goal attainment scaling (GAS) in rehabilitation: a practical guide. *Clinical rehabilitation* **23**(4): 362-370, DOI: 10.1177/0269215508101742.
Turner-Stokes L, Baguley IJ, De Graaff S, Katrak P, Davies L, McCrory P, Hughes A. (2010). Goal attainment scaling in the evaluation of treatment of upper limb spasticity with botulinum toxin: a secondary analysis from a double-blind placebo-controlled randomized clinical trial. *Journal of rehabilitation medicine* **42**(1): 81-89, DOI: 10.2340/16501977-0474.
Van Cleave JH, Egleston BL, Bourbonniere M, McCorkle R. (2011). Combining Extant Datasets with Differing Outcome Measures Across Studies of Older Adults after Cancer Surgery. *Research in Gerontological Nursing* **4**(1): 36-46, DOI: 10.3928/19404921-20101201-02.
Ware JE, Jr., Sherbourne CD. (1992). The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* **30**(6): 473-483.

Williams J, Leamy M, Bird V, Le Boutillier C, Norton S, Pesola F, Slade M. (in press). Development and evaluation of a measure to identify mental health service support for recovery (INSPIRE). *Social Psychiatry and Psychiatric Epidemiology*
Wise EA. (2004). Methods for analyzing psychotherapy outcomes: a review of clinical significance, reliable change, and recommendations for future directions. *Journal of personality assessment* **82**(1): 50-59, DOI: 10.1207/s15327752jpa8201_10.
Zahra D, Hedge C. (2010). The reliable change index: Why isn't it more popular in academic psychology? *Psychology Postgraduate Affairs Group Quarterly* **76**: 14-19.

Table 1. Sociodemographic characteristics of participants

		Pilot study	Feasibility study	RCT evaluation
N (complete data)		20	75	340
Age (years)		40.9 years	42.3 (10.7)	43.5 (10.8)
Gender				
	Male	14 (70)	51 (68)	214 (63)
	Female	6 (30)	24 (32)	125 (37)
Ethnicity				
	White	10 (50)	35 (47)	203 (60)
	Black	6 (30)	32 (43)	85 (25)
	Other	4 (20)	8 (11)	50 (15)
Diagnosis				
	Psychosis			340 (100)
	Schizophrenia	5 (26)	35 (50)	
	Bipolar disorder	5 (26)	15 (21)	
	Depression	1 (5)	1 (1)	
	Mixed	4 (21)	14 (20)	
	Other	4 (21)	5 (7)	

Table 2. PPO domains and baseline completion rates for Feasibility study (n=75) and evaluation in an RCT 3 (n=340)

Outcome domain descriptions given to participants	Outcome domain being described	Associated measure	Feasibility study	Evaluation in an RCT	
			N (%)	N (%)	Raw mean (SD)
1. Feeling more hopeful about the future	Hope	Herth Hope Index	7 (9)	47 (14)	36.14 (5.1)
2. Feeling more in control of my life	Empowerment	Mental Health Confidence Scale	6 (8)	26 (8)	66.95 (17.1)
3. Feeling more positive about myself	Self-esteem	Rosenberg Self-Esteem Scale	11 (15)	61 (18)	26.32 (4.4)
4. Feeling better treated by other people	Stigma	Stigma Scale	2 (3)	7 (2)	51.27 (13.2)
5. Feeling like you have meaning in your life	Meaning of life	Meaning of Life Questionnaire	7 (9)	42 (12)	24.45 (7.1)
6. Feeling supported by other people	Social support	MOS Social Support Survey	3 (4)	8 (2)	71.53 (13.9)
7. Feeling part of the community	Community integration	Community Integration Measure	3 (4)	11 (3)	30.46 (8.1)
8. Feeling better about your life	Wellbeing	Warwick-Edinburgh Mental Well-being Scale	8 (11)	39 (12)	45.26 (9.7)
9. Being able to manage day-to-day life	Daily functioning	Service user Perception of Functioning Scale	11 (15)	28 (8)	16.71 (3.6)
10. Having a better quality of life	Quality of life	Manchester Short Assessment of Quality of Life	17 (23)	71 (21)	55.90 (10.2)

Table 3. Reliable Change Index at one-year RCT follow-up (n=239)

	Cut-off (RCI=1.96)	Change status n (%)		
		Declined	No change	Improved
HHI	6.78	2 (6)	26 (79)	5 (15)
MHCS	8.69	5 (26)	8 (42)	6 (32)
RSES	7.30	1 (2)	42 (89)	4 (9)
Stigma	15.39	0 (0)	3 (100)	0 (0)
MLQ presence	8.34	1 (4)	24 (86)	3 (11)
MOS	11.62	2 (33)	4 (67)	0 (0)
CIM	7.70	0 (0)	3 (75)	1 (25)
WEMWBS	8.64	3 (12)	21 (84)	1 (4)
PPFS	2.36	2 (9)	14 (64)	6 (27)
MANSA	11.03	4 (8)	44 (85)	4 (8)
	Total	**20 (8)**	**189 (79)**	**30 (13)**