



Patient centric measures for a patient centric era: Agreement and convergent between ratings on The Patient Global Impression of Improvement (PGI-I) scale and the Clinical Global Impressions – Improvement (CGI-S) scale in bipolar and major depressive disorder

M. Mohebbi^{a,b,*}, S. Dodd^{b,d,e}, O.M. Dean^{b,c,d}, M. Berk^{b,c,d,e}

^a Deakin University, Faculty of Health, Biostatistics Unit, Geelong, Victoria, Australia

^b Deakin University, School of Medicine, IMPACT Strategic Research Centre, Barwon Health, Geelong, Victoria, Australia

^c Florey Institute for Neuroscience and Mental Health, University of Melbourne, Kenneth Myer Building, Parkville, Australia

^d University of Melbourne, Department of Psychiatry, Level 1 North, Main Block, Royal Melbourne Hospital, Parkville, Australia

^e Orygen, The Centre of Excellence in Youth Mental Health, The Department of Psychiatry and the Florey Institute of Neuroscience and Mental Health, The University of Melbourne, Australia

ARTICLE INFO

Article history:

Received 19 December 2017

Received in revised form 15 May 2018

Accepted 15 May 2018

Available online 30 May 2018

Keywords:

Patient global impression scale of improvement (PGI-I)
 Clinician global impression scale of improvement (CGI-I)
 Convergent validity
 Agreement
 Depression
 Bipolar disorder
 Psychology
 randomized control trial

ABSTRACT

Background: Concordant with an increased emphasis on consumer engagement, the Patient Global Impression Scale of Improvement (PGI-I) is commonly used as an outcome measure in studies evaluating the efficacy of treatments in medical and psychiatric conditions with subjective symptom domains. The current study evaluated the agreement between PGI-I and Clinician Global Impression Scale of Improvement (CGI-I) ratings and convergent validity of PGI-I among individuals with bipolar or major depressive disorders.

Method: Data were derived from three double-blind, placebo-controlled, multicentre studies conducted from 2007 to 2015 among adult individuals (N = 472). Clinicians were asked to rate participants symptoms using the CGI-I as well as severity of depression by the Montgomery-Åsberg Depression (MADRS), quality of life (Q-LES-Q), social and occupational functioning (SOFAS), and functional impairment (LIFE-RIFT). Participants were asked to assess their symptom improvement with the PGI-I. Bland-Altman agreement plots and Intra-class correlations were used to evaluate agreement, and Spearman correlation coefficients were implemented to examine convergent validity. Sub-group analyses for disorder type (bipolar and major depression) were performed.

Results: There was high agreement between the PGI-I and CGI-I ratings across follow-up time points (weeks 2, 4, 6, 8, 12, 16, 20, 24, and 28). Similar results were observed in male only and female only data and after adjustment for age and gender. Both PGI-I and CGI-I ratings were robustly positively correlated with MADRS, and LIFE-RIFT and negatively correlated with SOFAS and Q-LES-Q, supporting the convergent validity of the PGI-I. Sub-group analyses for bipolar and major depressive disorder showed similar findings.

Conclusion: Our findings support the utility of the PGI-I as a participant rated measure of global improvement among individuals with bipolar or major depressive disorders.

© 2018 Elsevier Masson SAS. All rights reserved.

Abbreviations: CGI-I, clinician global impression scale of improvement; DSM-IV, Diagnostic and Statistical Manual of Mental Disorders–Fourth Edition; ICH-GCP, International Conference of Harmonisation for Good Clinical Practice Guidelines; ICC, Intraclass Correlation Coefficient; MDD, major depressive disorder; MADRS, Montgomery-Asberg Depression; NAC, N-acetylcysteine; PGI-I, patient global impression scale of improvement; Q-LES-Q, Quality of Life Enjoyment and Satisfaction Questionnaire–Short Form; LIFE-RIFT, Range of Impaired Functioning Tool; SOFAS, Social and Occupational Functioning Scale.

* Corresponding author at: Deakin University, Faculty of Health, Biostatistics Unit, Geelong, Victoria, Australia.

E-mail address: m.mohebbi@deakin.edu.au (M. Mohebbi).

<http://dx.doi.org/10.1016/j.eurpsy.2018.05.006>

0924-9338/© 2018 Elsevier Masson SAS. All rights reserved.

1. Introduction

There is increasing emphasis on objective assessment of patient centric outcomes that span function, symptoms and quality of life. Validated measures of patient-reported outcomes with standardized questionnaires are thus critical to clinical and research outcome assessment [1]. When patients seek treatment, a determination of severity before treatment and improvement after treatment is mostly based on their subjective reporting of symptoms, and are linked to objective measures in those disorders

where they are salient. Little is known about how the patient's subjective and the clinician's objective rating of disorder improvement or treatment effect are aligned.

The Patient's Global Impressions of Improvement (PGI-I) scale has been included in several studies conducted worldwide to assess patients' overall perception of their condition by a simple and easy-to-use validated questionnaire [2]. The PGI-I is a 1-item questionnaires that ask an individual patient to rate the perceived change in his/her condition in response to therapy at endpoint. It is derived from the Clinical Global Impressions – Improvement scale (CGI-I) which was first developed for use in psychopharmacology trials as part of the NIMH collaborative study of schizophrenia [2]. Since then, it has been used as a standard primary outcome measure in studies investigating the efficacy of pharmacological treatments for psychiatric and medical conditions where subjective symptoms predominate, including pain, fatigue and mood [3–6] as well as secondary outcomes and responder analysis in many more studies; for example [7–9]. The CGI-I address the patient's improvement from baseline rated by the clinician. Both PGI-I and CGI-I show a bipolar scaling from 1 (very much improved) to 7 (very much worse). These types of measures have been validated in clinical studies of patients with stress incontinence [10], urogenital prolapse [11], fibromyalgia [12] major depressive disorder [13] and stress urinary incontinence [14].

This article aims to evaluate the agreement between patient- and clinician-rated global impression of improvement (PGI-I, CGI-I) scales. We also examine convergent validity of PGI-I compared with CGI-I correlation other clinician-assigned ratings of disease severity, functioning and quality of life. Data were derived from a three double-blind, placebo-controlled, multicentre, randomized controlled trials in adult outpatients with bipolar depression and major depressive disorder.

2. Methods

2.1. Study design and participants

This was a secondary analysis of data from 3 clinical trials. Details of the study designs and populations have previously been published [15–18]. Study 1 was a randomized, double-blind, placebo-controlled, parallel-design study to evaluate the efficacy of 2 g/day *N*-acetylcysteine (NAC) as adjunct maintenance treatment for bipolar disorder [15,16]. Participants ($n = 149$) had a Montgomery Asberg Depression Rating Score of (MADRS) ≥ 12 at trial entry and, after eight weeks of open-label NAC treatment, were randomized to adjunctive (in addition to treatment as usual) NAC or placebo for a further 24 weeks. Study participants were men and women residing in Australia and Brazil (www.anzctr.org.au: ACTRN12607000074493).

Study 2 was a randomized, double-blind, placebo-controlled, parallel-design study to evaluate the efficacy of 1 g/day NAC for major depressive disorder (MDD) in addition to existing treatments [17]. Participants ($N = 252$) had MADRS ≥ 18 at the time of entry with a current episode of MDD diagnosed according to DSM-IV-TR criteria. Participants were treated with NAC or placebo in addition to treatment as usual for 12 weeks and were followed to 16 weeks. Study participants were men and women residing in Australia (www.anzctr.org.au: ACTRN12607000134426).

Study 3 was a randomized, double-blind, placebo-controlled, parallel-design study to evaluate the efficacy of 200 mg/day of adjunctive minocycline or placebo for major depressive disorder (MDD) in addition to existing treatments [18]. Participants ($N = 71$) had MADRS ≥ 25 at the time of entry and met criteria for unipolar depression, based on Diagnostic and Statistical Manual of Mental Disorders–Fourth Edition (DSM-IV) criteria. Participants were randomized to NAC or placebo (parallel groups) over 12 weeks of

treatment and were followed to week 16. Study participants were men and women residing in Australia and Thailand (www.anzctr.org.au: ACTRN12612000283875).

2.2. Instruments

2.2.1. Patient Global Impression of Improvement (PGI-I) and Clinician Global Impression of Improvement (CGI-I)

Patient global impression of improvement scale (PGI-I) is a single-item global rating of change scale that ask an individual patient to rate the severity of a specific condition at baseline and or to rate at endpoints the perceived change in his/her condition in response to therapy. There are seven possible responses (scored 1–7): very much better, much better, a little better, no change, a little worse, much worse, and very much worse. The clinical global impression of improvement scale (CGI-I) is the clinician rated single-item scale that uses the same seven-point response criteria as the PGI-I [2] (see Appendix A in Supplementary material).

2.2.2. Depression severity

Severity of depressive symptomatology across studies time points were measured using the Montgomery–Åsberg Depression Rating Scale (MADRS) [19].

2.2.3. Quality of life

Quality of Life Enjoyment and Satisfaction Questionnaire–Short Form (Q-LES-Q) [20] was used for measuring quality of life.

2.2.4. Functional impairment

Functional impairment was measured using the Range of Impaired Functioning Tool (LIFE–RIFT) [21].

2.2.5. Social and occupational functioning

The Social and Occupational Functioning Scale (SOFAS) [22] was used to measure functioning over the duration of the study.

2.3. Ethics

All trials were conducted according to the Declaration of Helsinki 1964 as revised in 2008, the requirements of the Australian National Statement on Ethical Conduct in Human Research, the federal patient privacy (HIPAA) law and the International Conference of Harmonisation for Good Clinical Practice Guidelines (ICH-GCP) and were approved by institutional review boards at all sites.

2.4. Statistical analysis

Weighted agreement [23] was reported as a descriptive measure. The weights were given by $1 - \frac{|i-j|}{(k-1)}$, where i and j index the rows of columns of the ratings for CGI-I and PGI-I, $|i - j|$ indicate absolute difference and k is the maximum number of possible ratings. A weight of 1 indicates that an observation should count as perfect agreement and a weight of, say, 0.66 means that CGI-I and PGI-I are in two-thirds agreement (which happens if CGI-I and PGI-I are “two apart”). The agreements between clinician and patient ratings were assessed using Intraclass Correlation Coefficient (ICC) and its 95% confidence interval (CI) by implementing two-way random-effects model [24]. According to Fleiss [23], ICC values lower than 0.40 can be interpreted as poor, between 0.41 and 0.75 as fair, and above 0.75 as excellent agreement. The Bland-Altman plot [25] was used to visually inspect agreement. This analysis involved plotting the difference between CGI-I and PGI-I measurements against the average of the two measurements ± 1.96 times its SD known as the 95% limits of agreement.

For all analyses, participants were included regardless of what type of treatment they received. Convergent and divergent validity of PGI-I was evaluated by calculating Spearman correlations of PGI-I with CGI-I and four other clinician rated scale measuring depressive symptoms severity (MADRS), quality of life (Q-LES-Q), social and occupational functioning (SOFAS), and functional impairment (LIFE-RIFT). Direction and strength of CGI-I correlations with MADRS, Q-LES-Q, LIFE-RIFT and SOFAS were compared with PGI-I correlations with MADRS, Q-LES-Q, LIFE-RIFT and SOFAS, respectively. Convergent validity of PGI-I was evaluated with the expectation that PGI-I is positively correlated with MADRS and LIFE-RIFT and the strength of the correlations across follow-ups are similar to CGI-I correlations with MADRS and LIFE-RIFT. Furthermore convergent validity was examined with the expectation that SOFAS and Q-LES-Q are negatively correlated with PGI-I and strength of correlations across time points are similar between CGI-I and PGI-I. Pairwise correlations between CGI-I and PGI-I across follow-up time points are also reported as reference values for examining convergent validity. As CGI-I and PGI-I are measures of illness severity, convergent validity of PGI-I with change from baseline values of MADRS (MADRS change), LIFE-RIFT (LIFE-RIFT change), SOFAS (SOFAS change) and Q-LES-Q (Q-LES-Q change) are also evaluated in a similar manner. Sub-group analyses using similar analytical approaches were performed to evaluate agreement, convergent and divergent validity of PGI-I for both bipolar and MDD patients.

3. Results

A total of 472 individuals, (female 307, 65%) were randomized in the 3 studies; mean age ranged from 45.8 to 50.2 years. A total of 200 had MDD and 148 had bipolar depression (Table 1). Mean \pm SD of PCI-I and CGI-I values at each post-baseline assessment time point are presented in Table 2. There was a systematic decrease in both PCI-I and CGI-I values across time, reflecting clinical improvement with treatment. However, PCI-I and CGI-I mean values and SD were very similar at each time point. The weighed absolute agreement ranged from 94.27% to 98.69% showing a very high level of agreement. The unadjusted and adjusted (adjusted for gender and age) ICC of all time points were excellent.

Fig. 1 illustrates CGI-I and PGI-I Bland-Altman agreement plots across follow-up time points. The SDs of agreement from the

Bland-Altman analysis were as follows: week 2, ± 1.37 ; week 4, ± 1.91 ; week 6, ± 2.05 ; week 8, ± 2.13 , week 12, ± 2.21 , week 16, 2.29; week 20, ± 2.73 , week 24, ± 2.39 ; and week 28, ± 2.41 , showing an adverse association between time and agreement. The mean difference between CGI-I and PGI-I were close to zero across all follow-up time points in overall, male only and female only data (Table 3) illustrating negligible measurement bias between CGI-I and PGI-I. The direction of mean differences (CGI-I – PGI-I) were randomly changed across time points and in overall, male only and female only data showing no time trend in systematically higher or lower mean CGI-I values compared to mean PGI-I values or vice versa. The number of pairs outside 95% limits of agreement ranged from 2.27 to 8.99 showing reasonably acceptable out of range pairs of agreement all across time points except week 20. Comparing number of pairs above and below zero (x axes in Fig. 1 plots) and mean difference line (dashed lines in Fig. 1 plots) showed no tendency in higher or lower CGI-I rating compared to PGI-I rating or vice versa across time points. Fig. 1 also revealed acceptable homogeneity in CGI-I and PGI-I agreement across low, middle and high values of mean CGI-I and PGI-I values.

The convergent validity of the PGI-I was evaluated comparing MADRS and LIFE-RIFT pairwise correlations with CGI-I versus MADRS and LIFE-RIFT pairwise correlations with PGI-I across follow-ups (Table 4). As expected moderately high correlations were observed between CGI-I with MADRS and LIFE-RIFT, confirming high agreement between clinician assessed scales. Similar pairwise correlations were observed between PGI-I with MADRS and LIFE-RIFT, supporting convergent validity of PGI-I compared to clinician assessed scales. Convergent correlations were weaker than PGI-I and CGI-I correlations, showing higher agreement between patient and clinician measures of illness severity compared to other clinician assessed scales. Of note, both MADRS and LIFE-RIFT convergent values increased by follow-up time points (see Table 4). MADRS change from baseline also had moderately high correlations with both CGI-I and PGI-I, illustrating convergent validity of PGI-I. In a similar manner, convergent validity of PGI-I were examined by observing moderately high adverse correlation with similar strength between CGI-I with SOFAS and Q-LES-Q versus PGI-I with SOFAS and Q-LES-Q. Table S2 illustrates weighed absolute agreement, unadjusted and adjusted ICC of all time points between PGI-I and CGI-I stratified by mental disorder type (i.e. bipolar and MDD) and Table S3 shows the sub-group analyses findings for convergent and divergent

Table 1
Demographic and illness characteristics at baseline across 3 studies.

Variable	Test statistics	Value		
		Study 1 ^a	Study 2 ^b	Study 3 ^c
Age at entry to study	M (SD)	45.8 (11.4)	50.2 (12.7)	48.8 (14.7)
Gender% Female	% (n)	67.8 (101)	63.1 (159)	66.2 (47)
Diagnosis	% (n)			
Bipolar I disorder		69.6 (103)	–	–
Bipolar II disorder		29.7 (44)	–	–
Bipolar NOS		0.7 (1)	–	–
Major depressive disorder		–	100 (252)	100 (71)
Risk of suicide ^d % Yes	% (n)	70.9 (105)	75.5 (189)	59.2 (42)
Age of clinical diagnosis	M (SD)	35.9 (11.6)	35.7 (13.3)	36.2 (13.5)
Duration of illness since diagnosis (years)	M (SD)	10.0 (9.4)	14.4 (12.6)	14.0 (12.5)
Number of psychiatric hospitalisations	M (SD)	3.0 (4.4)	0.7 (1.9)	0.8 (1.7)
Smoker % Yes	% (n)	33.6 (50)	18.7 (47)	11.3 (8)
Alcohol abuse % Yes	% (n)	4.6 (7)	3.2 (8)	3.0 (2)
Substance use % Yes	% (n)	4.0 (6)	5.2 (13)	1.4 (1)
Anxiety ^e		51.7 (78)	69.0 (174)	57.7 (41)

^a Study 1: The efficacy of N-acetylcysteine as an adjunctive treatment in bipolar depression: An open label trial, ACTRN12607000074493.

^b Study 2: The Efficacy of Adjunctive N-Acetylcysteine in Major Depressive Disorder: A Double-Blind, Randomized, Placebo-Controlled Trial, ACTRN12607000134426.

^c Study 3: Adjunctive minocycline treatment for major depressive disorder: A proof of concept trial; ACTRN12612000283875.

^d Risk of suicide based on the Mini-International Neuropsychiatric Interview [30].

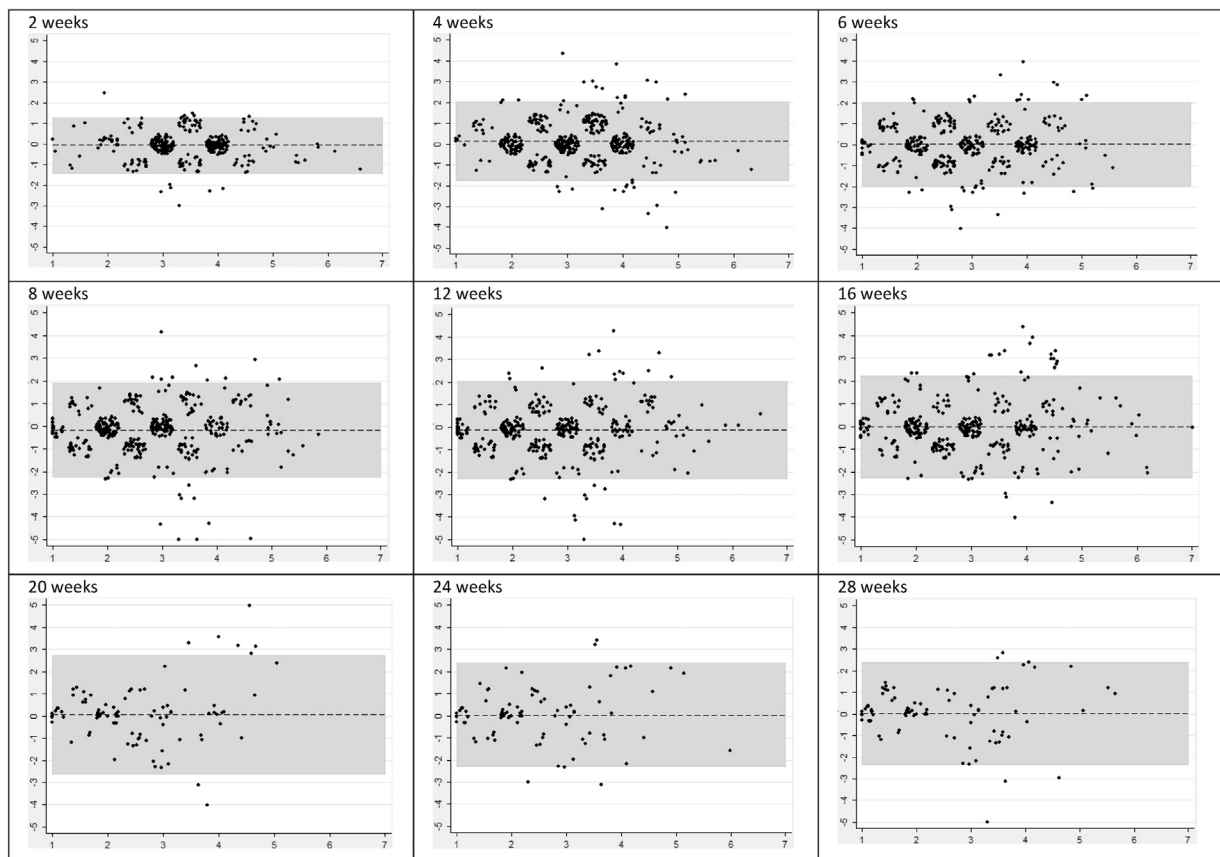
^e Pooled DSM-IV anxiety disorders.

Table 2

The agreement between PGI-I and CGI-I at all post baseline assessment time points.

Assessment time point	Number of pairs	CGI-I	PGI-I	Weighted agreement (%)	ICC unadjusted	ICC Adjusted
		Mean \pm SD	Mean \pm SD		ICC (95% CI)	ICC (95% CI)
2 weeks	308	3.40 \pm 0.90	3.46 \pm 0.92	98.69	0.93 (0.83, 0.98)	0.92 (0.81, 0.97)
4 weeks	437	3.24 \pm 1.00	3.10 \pm 1.05	97.41	0.89 (0.73, 0.96)	0.87 (0.69, 0.95)
6 weeks	347	2.90 \pm 1.12	2.88 \pm 1.09	95.80	0.87 (0.64, 0.96)	0.88 (0.65, 0.97)
8 weeks	391	2.72 \pm 1.16	2.88 \pm 1.15	96.80	0.85 (0.68, 0.93)	0.80 (0.61, 0.91)
12 weeks	365	2.74 \pm 1.27	2.87 \pm 1.22	96.55	0.85 (0.69, 0.93)	0.82 (0.65, 0.92)
16 weeks	353	2.93 \pm 1.31	2.93 \pm 1.24	96.37	0.83 (0.66, 0.92)	0.77 (0.60, 0.89)
20 weeks	89	2.62 \pm 1.34	2.55 \pm 1.14	94.88	0.81 (0.52, 0.95)	0.79 (0.47, 0.94)
24 weeks	83	2.53 \pm 1.29	2.49 \pm 1.20	96.08	0.87 (0.63, 0.96)	0.86 (0.60, 0.96)
28 weeks	88	2.43 \pm 1.33	2.41 \pm 1.32	94.27	0.85 (0.54, 0.96)	0.81 (0.48, 0.95)

Note: SD: Standard deviation. ICC: Intra-class correlation, CI: Confident interval.

**Fig. 1.** Bland and Altman plots at all post-baseline assessment time points to describe agreement between PGI-I and CGI-I.

Note. The y axes indicate difference between CGI-I and PGI-I. The x axes indicate the average of CGI-I and PGI-I. Shaded areas present agreement limits.

correlations of PGI-I with CGI-I, MADRS, LIFE-RIFT, SOFAS, and Q-LES-Q. Similar to the overall analysis, the weighed absolute agreements and ICCs confirmed the very high level of agreement between PGI-I and CGI-I across all time points. Similar correlation patterns were observed between PGI-I and CGI-I with MADRS, SOFAS, Q-LES-Q and LIFE-RIFT, confirming convergent and divergent validity of PGI-I in bipolar and MDD patients.

4. Discussion

The present study examined properties of patient's global impressions of improvement, an outcome measure commonly used in clinical trials for the treatment of medical and psychiatric disorders with subjective endpoints. Overall, our findings support

the utility of the PGI-I ratings among patients with MDD and bipolar disorder. These findings were replicated in bipolar and MDD patient sub-group analyses. PGI-I ratings were strongly associated with clinician's global impressions of improvement and other clinician-administered measures of specific symptomatology and quality of life and functioning.

The agreement between the CGI-I and the PGI-I was investigated by calculating Intra-class correlations and Bland and Altman plots. We found that estimation of improvement by the clinician and patient was excellent according to unadjusted ICCs, and very good according to ICCs that were adjusted to patients' age and gender. The findings was confirmed with close examination of Bland and Altman plots across gender and overall. Results of ICCs and Bland and Altman plots showed there was no gender bias in

Table 3

Bland and Altman plot statistics, including level of agreement and number of points outside the agreement limits intervals.

Assessment time point	Overall		Male		Female	
	Mean difference ^a (95% CI for LoA)	Outside LoA pairs; n (%)	Mean difference ^a (95% CI for LoA)	Outside LoA pairs; n (%)	Mean difference ^a (95% CI for LoA)	Outside LoA pairs; n (%)
2 weeks	-0.07 (-1.41, 1.27)	7/308 (2.27)	-0.02 (-1.26, 1.22)	1/111 (0.90)	-0.10 (-1.49, 1.30)	6/197 (3.05)
4 weeks	0.14 (-1.73, 2.02)	24/437 (5.49)	0.09 (-1.68, 1.87)	14/159 (8.81)	0.17 (-1.76, 2.10)	16/278 (5.76)
6 weeks	0.02 (-1.99, 2.03)	25/347 (7.20)	-0.03 (-2.01, 1.94)	6/127 (4.72)	0.05 (-1.99, 2.08)	13/220 (5.91)
8 weeks	-0.16 (-2.25, 1.92)	24/391 (6.14)	-0.15 (-2.09, 1.79)	9/135 (6.67)	-0.17 (-2.33, 1.99)	15/256 (5.86)
12 weeks	-0.13 (-2.30, 2.04)	16/365 (4.38)	-0.10 (-2.29, 2.09)	5/130 (3.85)	-0.15 (-2.31, 2.02)	11/235 (4.68)
16 weeks	0.003 (-2.24, 2.25)	18/353 (5.10)	-0.10 (-2.32, 2.11)	6/126 (4.76)	0.06 (-2.19, 2.32)	12/227 (5.29)
20 weeks	0.07 (-2.61, 2.74)	8/89 (8.99)	0.03 (-2.86, 2.92)	3/30 (10.00)	0.09 (-2.50, 2.67)	5/59 (8.47)
24 weeks	0.04 (-2.30, 2.38)	4/83 (4.82)	-0.07 (-2.57, 2.43)	1/28 (3.57)	0.09 (-2.18, 2.36)	3/55 (5.45)
28 weeks	0.02 (-2.34, 2.38)	5/88 (5.68)	-0.13 (-3.32, 3.06)	1/31 (3.23)	0.11 (-1.66, 1.87)	5/57 (8.77)

Note. LoA: Limits of agreement.

^a Difference between CGI-I and PGI-I.**Table 4**

Convergent and divergent validity of PGI-I at all follow-up time points by measuring pairwise correlations between CGI-I, MADRS, SOFA, Q-LES-Q and LIFE-RIFT.

Assessment time point	CGI-I								PGI-I								
	PGI-I	MADRS	SOFAS	Q-LES-Q	LIFE-RIFT	MADRS change	SOFAS change	Q-LES-Q change	LIFE-RIFT change	MADRS	SOFAS	Q-LES-Q	LIFE-RIFT	MADRS change	SOFAS change	Q-LES-Q change	LIFE-RIFT change
Week 2	0.67	0.58	-0.33	-0.29	0.43	0.60	-0.35	-0.37	0.42	0.45	-0.22	-0.34	0.36	0.49	-0.32	-0.39	0.40
Week 4	0.72	0.68	-0.45	-0.41	0.50	0.60	-0.46	-0.44	0.46	0.59	-0.39	-0.40	0.45	0.56	-0.41	-0.40	0.43
Week 6	0.68	0.77	-0.61	-0.61	0.67	0.76	-0.60	-0.54	0.61	0.57	-0.51	-0.52	0.62	0.59	-0.50	-0.47	0.61
Week 8	0.72	0.74	-0.54	-0.49	0.62	0.68	-0.60	-0.52	0.55	0.62	-0.48	-0.41	0.50	0.57	-0.48	-0.46	0.50
Week 12	0.76	0.76	-0.53	-0.45	0.62	0.74	-0.54	-0.52	0.57	0.67	-0.49	-0.43	0.56	0.66	-0.45	-0.52	0.53
Week 16	0.71	0.68	-0.55	-0.34	0.59	0.68	-0.52	-0.42	0.52	0.60	-0.49	-0.32	0.51	0.61	-0.45	-0.41	0.50
Week 20	0.70	0.77	-0.52	-0.63	0.70	0.49	-0.30	-0.61	0.55	0.58	-0.52	-0.61	0.63	0.46	-0.32	-0.62	0.45
Week 24	0.78	0.75	-0.32	-0.59	0.59	0.50	-0.12 ^a	-0.52	0.32	0.59	-0.35	-0.53	0.53	0.43	-0.13 ^a	-0.46	0.30
Week 28	0.82	0.72	-0.39	-0.59	0.64	0.51	-0.21 ^a	-0.50	0.44	0.64	-0.30	-0.51	0.51	0.44	-0.14 ^a	-0.52	0.44

^a Non-significant correlations (all other correlations are significant at the 0.01 level (2-tailed)).

PGI-I. Similar conclusion was drawn for patients' age by comparing unadjusted and age-adjusted ICCs. There was no systematic trend (bias) of over-estimation or under-estimation of PGI-I across follow-up times as examined by mean CGI-I and PGI-I and Bland and Altman plots. A commonly raised question regarding the PGI and related measures is that when patients assigned a rating of global change, their reliance on memory of baseline functioning might compromise the validity of the rating. We observed a slow declining trend in agreement through follow-up time as illustrated by increasing percentage of pairs outside agreement limits from Bland and Altman analytic results, and decreasing ICCs. Despite this trend, there were acceptable agreement through all follow-up time points confirming with strong relationship between PGI-I ratings and CGI-I ratings at mid- and post-treatment follow-ups supported the assumption that PGI-I ratings reflected actual changes in functioning from baseline.

Global Improvement ratings were reassuringly shown to be highly correlated with self-reported and clinician-assessed indices of depressive symptoms and impairment [26,27]. PGI-I convergent validity is evaluated by comparing PGI-I correlations with clinician-assessed indices with equivalent CGI-I correlations using the knowledge that there are high correlations between CGI-I and other clinician-assessed indices. Evidence of convergent validity of the PGI-I with other clinician's rated scales namely MADRS, SOFAS, Q-LES-Q and LIFE-RIFT were provided in this study indicated that the PGI-I instruments could be a valuable and useful tool for clinical studies and practice. Sub-group analyses confirmed both convergent and divergent validity of PGI-I in bipolar and MDD patients.

The present study has some limitations. First, there may be some inherent differences in bipolar and major depression that would influence these findings. However, this did not appear to be a significant factor in the current analyses. Secondly, that participants

were recruited from several countries, increasing the heterogeneity of the sample. Conversely, this may enhance generalizability of the congruence between patient and clinician ratings.

5. Conclusion

While the regulatory evaluation of treatment benefit in clinical trials may require multi-item instruments to fully describe the impact of treatment on various symptoms [28], the PGI-I can provide an overall patient centric appraisal of their own condition. This is concordant with the philosophy of patient centric care and research that is increasingly the touchstone of modern clinical care and assessment, where the patient should be the final judge of their care and clinical change [29]. The PGI is brief, simple and hence practical for clinical use by virtue of its simplicity of administration and interpretability.

Availability of data and materials

The datasets analysed during the current study are not publicly available due to human research ethics restrictions. Data are however available from the authors upon reasonable request and upon institutional human research ethics permission.

Competing interests

The authors declare that they have no competing interests.

Authors' contribution

All authors involved in designing the study. OMD prepared datasets for analysis.

MM performed the analysis, and was a major contributor in writing the manuscript. All authors read and approved the final manuscript.

Acknowledgement

The data used in these analyses was provided from clinical trials supported by the NHMRC, Australian Rotary Health, the Stanley Medical Research Institute, the Brain and Behavior Foundation and an Australasian Society for Bipolar and Depressive Disorders/Servier Grant. MB is supported by a NHMRC Senior Principal Research Fellowship (1059660). OMD is supported by a R.D. Wright Biomedical NHMRC Research Fellowship (1145634).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.eurpsy.2018.05.006>.

References

- [1] Rotenstein LS., Huckman RS, Wagle NW. Making patients and doctors happier—the potential of patient-reported outcomes. *N Engl J Med* 2017;377(14):1309–12.
- [2] Guy W. ECDEU assessment manual for psychopharmacology, vol. 76. US Department of Health, Education, and Welfare, Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration, National Institute of Mental Health, Psychopharmacology Research Branch, Division of Extramural Research Programs; 1976.
- [3] Hellerstein DJ, Yanowitch P, Rosenthal J, Samstag LW, Maurer M, Kasch K, et al. A randomized double-blind study of fluoxetine versus placebo in the treatment of dysthymia. *Am J Psychiatry* 1993;150: 1169–1169.
- [4] Barlow DH, Gorman JM, Shear MK, Woods SW. Cognitive-behavioral therapy, imipramine, or their combination for panic disorder: a randomized controlled trial. *JAMA* 2000;283(19):2529–36.
- [5] Brady K, Pearlstein T, Asnis GM, Baker D, Rothbaum B, Sikes CR, et al. Efficacy and safety of sertraline treatment of posttraumatic stress disorder: a randomized controlled trial. *JAMA* 2000;283(14):1837–44.
- [6] Liebowitz MR, Schneier F, Campeas R, Hollander E, Hatterer J, Fyer A, et al. Phenelzine vs atenolol in social phobia: a placebo-controlled comparison. *Arch Gen Psychiatry* 1992;49(4):290–300.
- [7] Papakostas GI, Fava M. Does the probability of receiving placebo influence clinical trial outcome? A meta-regression of double-blind, randomized clinical trials in MDD. *Eur Neuropsychopharmacol* 2009;19(1):34–40.
- [8] Busner J, Targum SD. The clinical global impressions scale: applying a research tool in clinical practice. *Psychiatry (Edgmont)* 2007;4(7):28.
- [9] Schneider LS, Dagerman K, Insel PS. Efficacy and adverse effects of atypical antipsychotics for dementia: meta-analysis of randomized, placebo-controlled trials. *Am J Geriatr Psychiatry* 2006;14(3):191–210.
- [10] Yalcin I, Bump RC. Validation of two global impression questionnaires for incontinence. *Am J Obstet Gynecol* 2003;189(1):98–101.
- [11] Srikrishna S, Robinson D, Cardozo L. Validation of the patient global impression of improvement (PGI-I) for urogenital prolapse. *Int Urogynecol J* 2010;21(5):523–8.
- [12] Arnold LM, Clauw D, Wang F, Ahl J, Gaynor PJ, Wohlreich MM. Flexible dosed duloxetine in the treatment of fibromyalgia: a randomized, double-blind, placebo-controlled trial. *J Rheumatol* 2010;37(12):2578–86.
- [13] Demyttenaere K, Desai D, Petit C, Croenlein J, Brecht S. Patient-assessed versus physician-assessed disease severity and outcome in patients with nonspecific pain associated with major depressive disorder. *Prim Care Companion J Clin Psychiatry* 2009;11(1):8.
- [14] Yalcin I, Viktrup L. Comparison of physician and patient assessments of incontinence severity and improvement. *Int Urogynecol J* 2007;18(11):1291–5.
- [15] Berk M, Dean OM, Cotton SM, Gama CS, Kapczinski F, Fernandes B, et al. Maintenance N-acetyl cysteine treatment for bipolar disorder: a double-blind randomized placebo controlled trial. *BMC Med* 2012;10(1):91.
- [16] Berk M, Dean O, Cotton SM, Gama CS, Kapczinski F, Fernandes BS, et al. The efficacy of N-acetylcysteine as an adjunctive treatment in bipolar depression: an open label trial. *J Affect Disord* 2011;135(1):389–94.
- [17] Berk M, Dean OM, Cotton SM, Jeavons S, Taniou M, Kohlmann K, et al. The efficacy of adjunctive N-acetylcysteine in major depressive disorder: a double-blind, randomized, placebo-controlled trial. *J Clin Psychiatry* 2014;75(6):628–36.
- [18] Dean OM, Kanchanatawan B, Ashton M, Mohebhi M, Ng CH, Maes M, et al. Adjunctive minocycline treatment for major depressive disorder: a proof of concept trial. *Aust N Z J Psychiatry* 2017 0004867417709357.
- [19] Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry* 1979;134(4):382–9.
- [20] Nee Endicott J, Harrison W, Blumenthal R. Quality of life enjoyment and satisfaction questionnaire: a new measure. *Psychopharmacol Bull* 1993;29(2):321–6.
- [21] Leon A, Solomon D, Mueller T, Turvey C, Endicott J, Keller M. The Range of Impaired Functioning Tool (LIFE–RIFT): a brief measure of functional impairment. *Psychol Med* 1999;29(4):869–78.
- [22] Goldman HH, Skodol AE, Lave TR. Revising axis V for DSM-IV: a review of measures of social functioning. *Am J Psychiatry* 1992;149(9):1148–56.
- [23] Fleiss JL, Levin B, Paik MC. Statistical methods for rates and proportions. John Wiley & Sons; 2013.
- [24] Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016;15(2):155–63.
- [25] Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res* 1999;8(2):135–60.
- [26] Khan A, Khan S, Shankles E, Polissar N. Relative sensitivity of the Montgomery–Asberg Depression Rating Scale, the Hamilton Depression rating scale and the Clinical Global Impressions rating scale in antidepressant clinical trials. *Int Clin Psychopharmacol* 2002;17(6):281–5.
- [27] Khan A, Brodhead AE, Kolts RL. Relative sensitivity of the Montgomery–Asberg depression rating scale, the Hamilton depression rating scale and the Clinical Global Impressions rating scale in antidepressant clinical trials: a replication analysis. *Int Clin Psychopharmacol* 2004;19(3):157–60.
- [28] Speight J, Barendse SM. FDA guidance on patient reported outcomes. *Br Med J (Online)* 2010;340:.
- [29] Stewart M. Towards a global definition of patient centred care: the patient should be the judge of patient centred care. *Br Med J* 2001;322(7284):444.
- [30] Hergueta T, Baker R, Dunbar GC. The Mini-International Neuropsychiatric Interview (MINI): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry* 1998;59(20):2233.