Psychiatry Research 208 (2013) 33–40

Contents lists available at SciVerse ScienceDirect

Psychiatry Research

journal homepage: www.elsevier.com/locate/psychres

Development and psychometric evaluation of the Discrimination and Stigma Scale (DISC)

Elaine Brohan [a,b,*], Sarah Clement [a], Diana Rose [a], Norman Sartorius [c], Mike Slade [a,1], Graham Thornicroft [a,1]

[a] King's College London, Health Service and Population Research Department, Institute of Psychiatry, London SE5 8AF, UK
[b] Adelphi Values Adelphi Mill, Grimshaw Lane, Bollington, Cheshire SK10 5JB, UK
[c] Association for the Improvement of Mental Health Programmes, 14 Chemin Colladon, 1209 Geneva, Switzerland

ARTICLE INFO

Article history:
Received 22 September 2012
Received in revised form
4 February 2013
Accepted 7 March 2013

Keywords:
Discrimination and Stigma Scale (DISC)
Psychometric validation
Scale development
Stigma
Mental illness

ABSTRACT

Mental illness is associated with unfair treatment in a number of areas of life. There is currently no psychometrically validated measure that has been developed to specifically focus on such experienced discrimination. This study aimed to finalise the Discrimination and Stigma Scale (DISC) and establish its psychometric properties. The DISC was further developed using (1) service user and interviewer focus groups; (2) reading ease testing; and (3) cognitive debriefing interviews. The revised scale then underwent psychometric testing to establish the following properties: reliability; validity; precision; acceptability; and feasibility. The final 22-item DISC demonstrated good psychometric properties ($n=86$) including inter-rater reliability (weighted kappa range: 0.62–0.95), internal consistency ($\alpha=0.78$) and test–retest reliability ($n=46$) (weighted kappa range: 0.56–0.89). Feasibility, validity and acceptability were also established. In conclusion, the 22-item DISC is recommended for use in measuring experienced stigma and discrimination. Additional work to develop a measure of anticipated stigma is recommended.

1. Introduction

Stigma is defined as a characteristic which individuals possess (or are believed to possess) that conveys an identity which is devalued in a particular social context (Crocker et al., 1998). Mental illness is associated with devaluation in a number of social contexts including: the workplace; healthcare settings; acting as a parent; and personal relationships (King et al., 2007; Link et al., 1997; Ritsher et al., 2003; Wahl, 1999). Three elements of stigma can be considered: (1) perceived stigma or the belief that the public hold negative attitudes towards people with a mental health problem, (2) experienced stigma or reported instances of unfair treatment or discrimination due to having a mental health problem and (3) self-stigma or adopting a stigmatised view of oneself (Yanos et al., 2008). Stigma research has largely focused on the measurement of perceived stigma. A recent review of studies using survey-based measures ($n=52$) reported that 79% used a measure of perceived stigma, 46% a measure of experienced stigma and 33% a measure of self-stigma (Brohan et al., 2010b). This review further suggests that, although several survey measures have addressed aspects of experienced stigma, there is currently no psychometrically validated measure that was developed with a specific focus on understanding the scope and content of these experiences.

The Discrimination and Stigma Scale (DISC) was developed to address this gap. It is based on the definition of Thornicroft et al. (2007), who present stigma as an overarching term including three elements: (1) problems of knowledge (ignorance or misinformation); (2) problems of attitudes (prejudice); and (3) problems of behaviour (discrimination) (Thornicroft et al., 2007). This definition moves the stigma focus from the characteristics of the individual to the problems that perpetuate stigma. The DISC focuses on the third aspect of stigma: problems of behaviour or discrimination. It is an interview-based measure which collects qualitative and quantitative data to provide a rating of the degree to which discrimination has been experienced in various areas of life including work, relationships, parenting, housing, leisure, and religious activities. The qualitative aspect of the scale asks respondents, for each domain, to give an example of how they have been treated differently (or not) from other people because of their diagnosis of mental illness. Participants are then asked to given a Likert scale rating for each item, which is the quantitative aspect of the scale.

* Corresponding author at: King's College London, Health Service and Population Research Department, Institute of Psychiatry, London SE5 8AF, UK.
Tel.: +44 0207 848 0765.
E-mail addresses: elaine.brohan@kcl.ac.uk, elainebrohan@yahoo.co.uk (E. Brohan).
[1] MS and GT contributed equally to this manuscript.

0165-1781/$ - see front matter
http://dx.doi.org/10.1016/j.psychres.2013.03.007

A preliminary version of the DISC was developed as part of the International Study on Discrimination and Stigma Outcomes (INDIGO) (Thornicroft et al., 2009). In scale development, face and content validity were established through a literature review, Delphi consultation and pilot testing of the draft scale were carried out within research teams at 28 participant study sites in 27 countries. Twenty-five interviews were conducted at each site (total $n=732$) with five of the interviews at each site audio-taped, transcribed verbatim, translated into English and qualitatively analysed by members of the study team (Rose et al., 2011). The results of this study suggested that negative discrimination was frequently experienced. For example, 344 (47%) reported discrimination in making or keeping friends, 315 (43%) in relationships with family members, and 209 (29%) in finding a job. Discrimination was also frequently anticipated, with 469 (64%) inhibiting themselves from applying for work, training, or education and 402 (55%) stopping themselves from looking for a close relationship for this reason. This initial version of the scale also provided the opportunity for participants to record any experiences of positive discrimination (i.e. situations where they had been treated more positively because of their mental health problem). Reports of positive experienced discrimination were rare. The qualitative and quantitative analysis of the data collected suggested that the scale may benefit from further developmental work to improve the relevance and ease of use of items and response options. This article details this work and further work to establish the psychometric properties of the revised scale.

1.1. Aims

This study aims to

(1) complete the developmental work to maximise the acceptability and feasibility of the DISC, leading to a finalised version of the scale (Phase 1); and
(2) establish the psychometric properties of the revised DISC (Phase 2).

2. Methods

2.1. Methods for completing the development of the DISC

Firstly, the corrections indicated by the INDIGO data analyses were incorporated, as described elsewhere (Rose et al., 2011; Thornicroft et al., 2009). As mentioned in Section 1, this focus improved the relevance and ease of use of items and response options. This included restructuring the response options from a 7-point Likert scale that allowed a rating of positive or negative discrimination on each items to a 4-point scale that allowed ratings of negative discrimination only. An additional positive discrimination subscale was introduced to allow participants to provide ratings on this aspect for several areas of life in which it was most frequently reported.

This was followed by three stages of evidence gathering to support further development: (1) service user and interviewer focus groups ($n=4$ groups); (2) reading ease testing; and (3) cognitive debriefing. Evidence from stage 1 was used to create a draft DISC, which was considered in stages 2–3.

2.1.1. Stage 1: service user and interviewer focus groups

Two semi-structured focus groups were conducted with mental health service users. Eligible participants were identified by the clinical team at a day centre service. Participants were asked to complete the DISC at the beginning of the focus group and then discuss aspects of the scale including: overall ease of completion; relevance of items and response options, time taken to complete and recommendations for improvement. Two focus groups were also separately conducted with interviewers who had experience in using the DISC as part of other studies run by colleagues at the Institute of Psychiatry. Group discussion focused on experience of using the scale and recommendations for improvement.

2.1.2. Stage 2: reading ease testing

The Flesch Reading Ease score and Flesch–Kincaid Grade level were assessed using MS Word. These are widely used tools which assess readability based on the syllabic and sentence structure of the text (Kalk and Pothier, 2008). The Flesch Reading Ease score ranges from 0 to 100 with higher scores being easier to read. The Flesch–Kincaid Grade level provides an indication of the US educational grade to which the material is most appropriate (range 0–17) (Flesch, 1974).

2.1.3. Stage 3: cognitive debriefing

Cognitive debriefing involves a small interview study, providing qualitative data on the mental processes that respondents use to answer questions (Ojanen and Gogates, 2006). Once changes had been made following the recommendations of stages 1 and 2, cognitive debriefing interviews were conducted as a check to ensure conceptual clarity and ease of meaning in the final DISC. Five individuals were recruited from a day centre service using the methods described in the service user focus groups.

2.2. Methods for the psychometric evaluation of the DISC

2.2.1. Design

A cross-sectional study design was used, with participants being interviewed at one point in time. A sub-sample of participants also completed the DISC again 7–14 days following initial administration to establish the test–retest reliability. This study, and the earlier work described in stages 1–3 above, received National Research Ethics Service (NRES) approval from the Camden and Islington Community Local Research Ethics Committee (REC ref: 08/H0722/40).

2.2.2. Sample

A sample size of 90 was chosen as it is sufficient to establish that the inter-rater reliability is at least 0.7 (assuming that the true level is 0.8) (Walter et al., 1998). Sample size was calculated based on inter-rater reliability as this was the property which required the largest sample. The sample of 90 was also sufficient to establish the other properties under consideration (e.g. convergent and discrimination validity analyses). Convenience sampling was used. All participants were aged 18 years or over and had used a community mental health team in the past 12 months.

2.2.3. Procedure

Interviews were conducted by a team of six researchers. Each researcher participated in a half-day training session led by EB. During the participant interviews, DISC responses were digitally recorded. A second member of the research team listened to the interview and recorded a score for each DISC item. This was compared with the original interview scores to calculate inter-rater reliability. Every seond participant in this study was asked to repeat the DISC interview again 7–14 days following initial administration. The study measures are detailed below.

2.2.4. Measures

2.2.4.1. Discrimination and Stigma Scale (DISC). The revised DISC is a 35-item, interview-based, measure. All items are scored on a 4-point Likert scale ranging from 0=not at all to 3=a lot. It comprises a global scale and four subscales, each of which is scored separately. The four subscales are: (1) Unfair treatment (22 items); (2) Stopping self (4 items); (3) Overcoming stigma (2 items) and (4) Positive treatment (7 items). The qualitative aspects of the scale were retained, with respondents asked, for each item, to give an example of how they have been treated differently (or not) from other people because of their diagnosis of mental illness. Both a mean and a total score are calculated for each subscale and the global scale. This allows both the typical level of stigma in each applicable area of life, and its spread over the different areas to be encompassed.

2.2.4.2. The Stigma Scale (SS). The SS is a 28-item self-completed measure. It has three subscales: disclosure (11 items), discrimination (12 items) and positive aspects (5 items) (King et al., 2007). The scale has good test–retest reliability (kappa range 0.49–0.71) and internal consistency ($\alpha=0.87$).

2.2.4.3. Internalised Stigma of Mental Illness Scale (ISMI). The ISMI is a 29-item self-completed measure that assesses mental health service users' experience of internalised stigma (Ritsher et al., 2003). It is composed of five subscales: Alienation, Stereotype Endorsement, Perceived Discrimination, Social Withdrawal and Stigma Resistance. Strong internal consistency ($\alpha=0.90$) and test–retest reliability ($r=0.92$) have been reported.

2.2.4.4. Brief Psychiatric Rating Scale (BPRS). The BPRS measures psychiatric symptomatology, including positive symptoms, general psychopathology and affective symptoms (Overall and Gorham, 1962). The 18-item version of the scale was used in this study (Lukoff et al., 1986). Items 1–10 are rated by the participant during an interview, while items 11–18 are rated by the researcher following observation of

the participant. Each item is rated on a 7-point scale anchored at 1=not present and 7=extremely severe.

2.2.4.5. Global Assessment of Functioning (GAF). The GAF comprises two numeric scales (Endicott et al., 1976). Each is rated by an observer from 0 (most severe) to 100 (least severe). One (GAF-D) asks the rater to consider the level of disability, and the other (GAF-S) symptoms. The GAF has been widely used and is a reliable and valid measure of functioning (Jones et al., 1995).

2.3. Data analysis

2.3.1. Completing the development of the DISC

Verbatim transcripts from the four focus groups were analysed for recurrent themes using the methods of content and thematic analysis (Joffe and Yardley, 2004). Suggestions of changes were grouped under these themes and discussed within the study team. The reading ease scores and cognitive debriefing results were also considered. This work focused on maximising the acceptability and feasibility of the DISC. Acceptability describes the extent to which a scale is targeted towards the intended population. The suitability of the wording of the survey to the target audience was established using the service user focus groups, service user research report, reading ease testing and cognitive debriefing interviews. Acceptability can also be used to describe the quality of data as assessed by the completeness of data and score distributions (Lamping et al., 2002). This property will be established as part of the psychometric evaluation of the DISC in Section 3.3.2. Feasibility asks whether the scale is easy to administer and process (Fitzpatrick et al., 1998). This was considered by conducting the interviewer focus groups, where individuals with experience of using DISCv11 were asked to discuss aspects related to feasibility. Further aspects of feasibility are also considered as part of the psychometric evaluation in Section 3.3.2.

2.3.2. Psychometric evaluation of the DISC

Analysis was performed using SPSS version 15 (SPSS Inc., 2006) and Stata version 9.2 (StataCorp, 2005). To determine whether the variables to be used in the analysis were sufficiently normally distributed, histograms were examined for all continuous variables. Skew and kurtosis values were examined and a Kolmogorov–Smirnov test was performed. Outliers were identified using the z-score criterion of ± 3.29, with those cases violating this threshold excluded from analysis. Descriptive statistics for the socio-demographic and illness-related variables were calculated.

2.3.2.1. Scoring. Descriptive statistics were calculated for all DISC scale and subscale scores. Positive treatment is reverse coded when treated as a subscale score so a high score on this variable indicates a lack of positive treatment or occasions when the person received help or support not available to others, as a result of the persons' mental health problem. A strategy for the interpretation of scores was applied using the DISC midpoint of 1.5. In this interpretation four categories were used to understand DISC mean scores: <1 minimal discrimination; 1–1.5 low discrimination; 1.5–2 moderate discrimination; and 2+ high discrimination. This scoring strategy was based on a previous approach in studies using the Internalized Stigma of Mental Illness (ISMI) scale, which represented a high level of self-stigma as an average score above the midpoint of 2.5 (Brohan et al., 2010a, 2011).

2.3.2.2. Reliability. The reliability of the DISC was assessed by considering: (1) consistency over sub-components (internal consistency); (2) consistency over raters (inter-rater reliability) and (3) consistency over time (test–retest reliability). Internal consistency was assessed using Cronbach's α with a criterion of $\alpha \geq 0.70$ indicative of appropriate internal consistency for each subscale (Cronbach, 1951). A criterion value over 0.90 was also flagged, as this may indicate item redundancy. Lin's concordance statistic (ρ_c) was used to calculate the overall inter-rater reliability and test–retest reliability for each subscale and the total DISC score (Lin, 1989). This was performed using the 'concord' command in Stata. Key demographic variables were examined to test for differences between those who were included in the retest study and those not included in the retest study. Due to the normal distribution, the t-test was conducted in the case of continuous variables (age, social contact) and the chi-square test or Fisher's exact test was used in the case of categorical data (all other variables). A criterion of Lin's $\rho_c \geq 0.70$ was used to indicate acceptable reliability for all scores. A weighted kappa score was also calculated for each item pair to provide details on the test–retest and inter-rater reliability of individual items. A criterion of weighted kappa ≥ 0.4 was used to indicate acceptable item level reliability.

2.3.2.3. Validity. The following aspects of validity were established in this study: (1) within-scale validity; and (2) convergent and discriminant validity. Within-scale analysis of validity provides evidence that a single construct is being measured and that items can be combined to form a total score. This is assessed based on internal consistency, as established in the reliability analysis. The correlations between scales were also examined to assess the degree to which they measure related aspects of the construct. Moderate correlations were taken as the criterion for this (Lamping et al., 2002). Convergent and discriminant validity are specific forms of construct validity where it is hypothesised that the scale under consideration will have a stronger association with certain variables (i.e. convergent validity) and a weaker association with other variables (i.e. discriminant) validity. The convergent validity of DISC subscales was examined as specified below. A significant moderate to strong correlation (0.3 or greater) was taken as the criterion for convergent validity. Terwee et al. (2007) propose that construct validity is established if at least 75% of results are in accordance with the hypothesised relationships. For the DISC, the following relationships are hypothesised:

1. the DISC Unfair Treatment subscale will have a significant association with the Discrimination subscale of the SS and the Discrimination Experiences subscale of the ISMI;
2. the DISC Stopping Self subscale will have a significant association with the Disclosure subscale of the SS and the Social Withdrawal subscale of the ISMI;
3. the DISC Overcome Stigma subscale will have a significant association with the Positive Aspects subscales of the SS and the Stigma Resistance subscale of the ISMI;
4. the DISC Positive Discrimination subscale will have a significant association with the Positive Aspects subscales of the SS and the Stigma Resistance subscale of the ISMI; and
5. the DISC Total Score will have a significant association with the SS Total score and SS Discrimination score.

Divergent validity was measured by assessing the relationship between DISC subscales and Total score with the demographic variable of gender. There is currently a lack of evidence on the relationship between gender and discrimination directed toward those with mental health problems (Angermeyer and Dietrich, 2006). This variable was selected as a measure of divergent validity as it is expected that DISC scores will not have a significant association with gender, as established in the previous INDIGO study (Thornicroft et al., 2009). No significant difference in DISC scores by gender will be taken as the appropriate criterion.

2.3.2.4. Precision. This considers the appropriateness of the scaling assumptions i.e. how well each item fits within its proposed scale. Corrected item-total correlations <0.30 were used to indicate unacceptable fit of the items with the scale total score (Terwee et al., 2007). Each item was also correlated with its own scale total and with other subscale totals. To maintain the precision of the subscales, each item needed to correlate more highly with its own subscale than with other DISC subscales.

2.3.2.5. Acceptability. Acceptability describes the quality of data as assessed by the completeness of data and score distributions (Lamping et al., 2002). The following aspects of acceptability were established: (1) Maximum endorsement frequencies (MEF); and (2) Aggregate adjacent endorsement frequencies (AEF). To consider MEF, the n (%) of respondents who endorse each response category for each item was presented. MEF $>80\%$ in a particular category indicates that the item may need further consideration. This includes floor and ceiling effects which are a specific instance of MEF violation at the lowest and highest scale points respectively. The AEF criterion is violated when two or more adjacent scale points on an item show $<10\%$ of the responses (The Whoqol Group, 1998).

2.3.2.6. Feasibility. Feasibility was considered by noting the time taken to conduct the DISC interview. A time of greater than 30 min was considered indicative of an overly long interview-based measure.

3. Results

3.1. Results of completing the development of the DISC

There were 13 participants in the service user focus groups and 12 in the interviewer focus groups. Males accounted for 39% of participants.The mean age of the participants was 46.8 (S.D.=13.42). Eighty percent reported some personal experience of stigma. Participants in the service user focus groups most commonly reported the following diagnoses: depression ($n=3$); schizophrenia ($n=4$); and bipolar disorder ($n=6$). In the interviewer focus groups participants had conducted there had been an average of 5.67 DISC interviews (S.D.=4.62). The recurrent themes identified by analysing the focus group transcripts suggested five areas of change: (1) reduction of complexity; (2) change phrase 'treated differently from other people'; (3) change phrase 'your diagnosis of mental illness'; (4) change response options; and (5) wording of individual questions. After these changes were incorporated, and the Flesch-Kincaid Grade

level was reduced from 13.2 to 7.4 and Flesch Reading Ease score increased from 36.2 to 65.7, suggesting that the modifications were successful in reducing the complexity of the scale. The cognitive debriefing interviews further support acceptability of the revised DISC. All items were well understood by the five participants and no further areas of clarification were identified. The finalised DISC is a 35-item scale as described in the methods for psychometric evaluation of the DISC.

3.2. Results of the psychometric evaluation of the DISC

3.2.1. Demographic characteristics

Eighty-six people took part in the study. Their socio-demographic and clinical characteristics are reported in Table 1. The BPRS mean score indicates that participants can be classified as moderately ill (Leucht et al., 2005). Of those included in this analysis, 39.8% (n=83) fall into this category with a score between 41 and 52; 24.1% were categorised as mildly ill with scores of 31–40; 19.3% were categorised as markedly ill with scores of 53 or greater; and the remaining 16.9% were classified as borderline ill with scores of 30 or less. This is in keeping with the reported GAF scores which suggest that the participants have a moderate level of symptoms and moderate difficulty in social or occupational functioning (Endicott et al., 1976).

Fig. 1 shows the frequency of endorsement for each DISC item.

Table 2 reports on the stigma measures. For the DISC, all subscale scores fell in the low stigma category (scores of 1–1.5), with the exception of Unfair Treatment, which was in the minimal stigma category, and Positive Treatment, which was in the high stigma category. On average participants reported experiencing stigma in 16.55 of the 35 aspects of life on the DISC. All but one of the participant endorsed at least one aspect in life in which they had experienced stigma (98.8%). The ISMI scores displayed are in

Table 1
Socio-demographic and clinical characteristics.

Variable		*N*	%
Gender	Male	50	58.1
	Female	35	40.7
Age (n=86)	Mean (S.D.)=41.2 (10.9)	Min=19, max=67	
Ethnicity	White British	32	37.2
	Black British	16	18.6
	Black African	12	14.0
	Black Caribbean	11	12.8
	Other	12	14.0
Highest level of education	Higher diploma or degree	33	38.4
	Vocational qualification	18	20.9
	GCSE/O level/CSE	13	15.1
	No formal qualifications	12	14
	A levels	8	9.3
Employment status	Unemployed	39	45.4
	Volunteer	21	24.4
	Student	14	16.3
	Work full time/part-time	12	14.0
	Retired	4	4.7
Relationship status	Single	48	55.8
	Girlfriend/boyfriend/partner	21	24.4
	Divorced or separated	14	16.3
	Married/ cohabiting	6	7.0
Self-reported diagnosis	Schizophrenia	26	30.2
	Bipolar disorder	24	27.9
	Depression	15	17.4
	Psychosis	11	12.8
	Do not know	8	9.3
	Anxiety disorder	4	4.7
	Personality disorder	3	3.5
	Other	6	7.0
Years since first treatment for MHP (n=83)	Mean (S.D.)=15.8 (10.5)	Min=1, max=45	
Ever admitted to hospital voluntarily?	Yes	69	82.1
Ever admitted to hospital under MHA?	Yes	43	50.0
No. of voluntary admissions to hospital (n=63)	Mean (S.D.)=3.4 (2.6)	Min=1, max=10	
No. of admissions to hospital under MHA (n=39)	Mean (S.D.)=2.6 (1.6)	Min=1, max=8	
BPRS total (n=83)	43.1 (12.0)	20–90	
Thought disturbance (n=84)	8.1 (4.3)	4–27	
Withdrawal (n=83)	8.8 (3.9)	4–21	
Anxiety–depression (n=85)	14.0 (6.2)	4–27	
Hostility suspiciousness (n=84)	6.5 (2.7)	3–13	
Activity (n=84)	5.8 (2.3)	3–14	
GAF disability (n=84)	60.4 (13.5)	35–90	
GAF symptoms (n=84)	59.6 (10.8)	31–81	

MHA=mental health act; MHP=mental health problem.

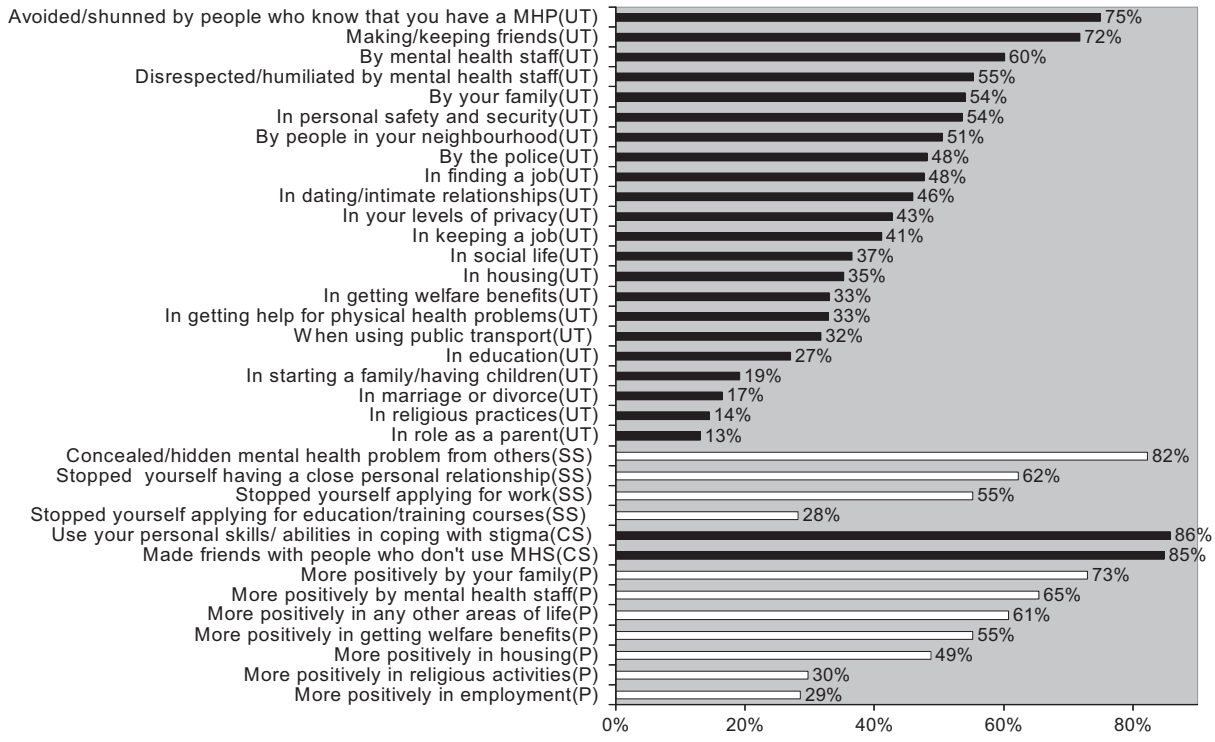

Fig. 1. Proportion of agree responses for DISC item. UT=unfair treatment subscale, SS=stopping self subscale, CS=stigma coping subscale, P=positive treatment subscale, MHP=mental health problem, MHS=mental health services.

Table 2
Stigma scores ($n=86$).

Variable	Mean (S.D.)	Min–max
DISC Total[a] ($n=84$)	1.3 (0.4)	0.4–2.6
DISC Unfair Treatment ($n=85$)	0.9 (0.6)	0–2
DISC Stopping Self ($n=84$)	1.3 (0.8)	0–3
DISC Overcoming Stigma[b] ($n=84$)	1.2 (1.0)	0–6
DISC Positive Treatment ($n=85$)	2.5 (1.6)	0.5–10.5
DISC count[c] ($n=84$)	16.6 (5.9)	0–30
DISC Unfair Treatment ($n=85$)	9.0 (4.4)	0–18
DISC Stopping Self ($n=84$)	2.3 (1.3)	0–4
DISC Overcoming Stigma ($n=84$)	1.7 (0.5)	0–3
DISC Positive Treatment[b] ($n=85$)	3.5 (1.9)	0–7
ISMI Total ($n=84$)	2.3 (0.4)	1.3–3.3
ISMI Alienation ($n=84$)	2.5 (0.6)	1–4
ISMI Stereotype Endorsement ($n=85$)	2.0 (0.5)	1–3
ISMI Discrimination Experience ($n=85$)	2.4 (0.6)	1–4
ISMI Social Withdrawal ($n=84$)	2.4 (0.6)	1–4
ISMI Stigma Resistance ($n=85$)	2.1 (0.5)	1–4
SS Total ($n=84$)	61.0 (14.1)	24–97
SS Discrimination Experience ($n=85$)	30.0 (8.2)	12–50
SS Disclosure ($n=85$)	23.5 (7.4)	7–39
SS Positive Aspects ($n=84$)	7.4 (2.5)	3–16

[a] DISC Total is the mean DISC scale or subscale score.
[b] DISC z has been reverse coded so a higher score means less positive discrimination.
[c] DISC count is a count score of the number of items endorsed in the DISC scale or subscale.

keeping with published scores for a similar population; for example, Lysaker et al. (2007) report the following: Alienation 2.31 (S.D. 0.65); Stereotype endorsement 1.99 (S.D. 0.54); Discrimination experience 2.42 (S.D. 0.69); Social withdrawal 2.30 (S.D. 0.66); and Stigma resistance 2.17 (S.D. 0.52). The displayed SS scores were also in keeping with published scores, as reported by King et al. (2007): total SS 62.6 (S.D. 15.4); Discrimination 29.1 (S.D. 9.5); Disclosure 24.7 (S.D. 8.0); and Positive aspects 8.8 (S.D. 2.8).

The degree to which the criterion for each psychometric property was established is summarised in Table 3. Each property will now be discussed.

3.2.2. Reliability

3.2.2.1. Internal consistency. Cronbach's α for the DISC was 0.78, and removal of any item did not increase the α value to greater than 0.79. Sub-scale analysis showed Cronbach's α for Unfair Treatment was 0.82, for Stopping Self was 0.66 and for Positive Treatment was 0.67. Values were not increased by the deletion of any items.

3.2.2.2. Inter-rater reliability. Lin's $\rho_c=0.89$, $p<0.001$. For individual items the weighted kappa statistic ranged from 0.62 (dating or intimate relationships) to 0.97 (stopped self applying for education). The overall inter-rater reliability for the Unfair Treatment subscale Lin's $\rho_c=0.87$, $p<0.001$, for the Stopping Self subscale Lin's $\rho_c=0.91$, $p<0.001$, for Overcoming stigma Lin's $\rho_c=0.78$, $p<0.001$ and for Positive Treatment Lin's $\rho_c=0.91$, $p<0.001$.

3.2.2.3. Test–retest reliability. There were no significant differences between those who took part in the retest study and those who did not on key demographic variables, including gender, age, employment status, ethnicity and diagnosis. The DISC ($n=44$) had an overall test–retest reliability Lin's $\rho_c=0.88$, $p<0.001$. For individual items the weighted kappa statistic ranged from 0.45 (treated more positively by mental health staff) to 0.89 (treated

Table 3
Summary of psychometric properties of DISC (35-item version).

Psychometric property	Constituent parts	Criterion	Was criterion met?
Reliability	Internal consistency	Cronbach's α criterion of ≥0.70 or greater for each subscale	Yes
	Inter-rater reliability	Lin's ρ_c for each subscale and total score criterion of ≥0.70	Yes
	Test–retest reliability	Lin's ρ_c for each subscale and total score criterion of ≥0.70	Yes
Validity	Within-scale analyses	Correlations between subscales. Moderate correlations are taken as the criterion (0.3–0.50)	Partially. DISC Unfair Treatment and Stopping Self reached threshold, while Overcoming Stigma and Positive Treatment did not
	Convergent validity	75% of the hypothesised relationships (1–6) as specified for convergent and divergent validity should be met	No. 50% (4/8) of the relationships were met. See 1–6 for more details.
		1. DISC Unfair Treatment sig association with SS Discrimination and ISMI Discrimination Experiences	Yes
		2. DISC Stopping Self sig association with SS Disclosure and ISMI Social Withdrawal	No, although significantly correlated, neither association met the threshold
		3. Overcome Stigma sig association with SS Positive Aspects and ISMI Stigma Resistance	Partially. Both sets were sig correlated but only ISMI Stigma Resistance reached threshold
		4. Positive Discrimination sig association with SS Positive Aspects and ISMI Stigma Resistance	No. Neither set of variables were significantly correlated
		5. DISC Total Score sig association with SS Total score and SS Discrimination	Partially. Both sets were sig correlated but only SS Discrimination reached threshold
	Divergent validity	6. DISC scores will have no sig association with gender	Yes
Precision	Corrected item-total correlations	Corrected item-total correlations < 0.30 indicate unacceptable fit of the items with scale total score	Partially. All but one item met the criterion
	Intercorrelation of items and subscales	Each item should correlate more highly with its own subscale than with other DISC subscales	Partially. All but one item correlated more highly with its own subscale than another
Acceptability	MEF	MEF $> 80\%$ in a particular category indicates this criterion has been violated	Yes
	AEF	AEF criterion is violated when two or more adjacent scale points on an item show $< 10\%$ of responses	No. Four items violated this criterion. However, this violation is conceptually justified (see Section 4—discussion)
Feasibility	Time taken to complete DISC	An average time to complete DISC in excess of 30 min	Yes. Average time was 29.06 min

unfairly in keeping a job). The overall test–retest reliability for the unfair treatment subscale ($n=46$) Lin's $\rho_c=0.89$, $p<0.001$, for the Stopping Self subscale ($n=45$) Lin's $\rho_c=0.71$, $p<0.001$, for Overcoming Stigma ($n=45$) Lin's $\rho_c=0.56$, $p<0.001$ and for Positive Treatment ($n=45$) Lin's $\rho_c=0.50$, $p<0.001$.

3.2.3. Validity

3.2.3.1. Inter-correlation of DISC subscales. The DISC Total, Unfair Treatment and Stopping Self subscales reached the threshold moderate level (0.3), while the Overcoming Stigma (correlation with Stopping Self $r=0.20$, NS) and Positive Treatment subscales (correlation with Unfair Treatment $r=0.14$, NS, correlation with Stopping Self $r=-0.02$, NS) did not meet this criterion.

3.2.3.2. Convergent validity. The DISC Unfair Treatment subscale showed adequate convergent validity with the Discrimination subscale of the SS ($r=0.54$, $p<0.001$) and the Discrimination Experience subscale of the ISMI ($r=0.31$, $p<0.001$). The DISC Stopped Self subscale was significantly correlated with the Disclosure subscale of the SS ($r=0.25$, $p<0.05$) and the Social Withdrawal subscale of the ISMI ($r=0.23$, $p<0.05$). Neither of these correlations reached the threshold of 0.3. The DISC Overcome Stigma subscale was significantly correlated with the Positive Aspects subscale of the SS ($r=-0.24$, $p<0.05$) and the Stigma Resistance subscale of the ISMI ($r=0.29$, $p<0.001$). Neither met the threshold of 0.3. The DISC Positive Treatment subscale was not significantly correlated with the Positive Aspects subscales of the SS ($r=-20$, NS) or the Stigma Resistance subscale of the ISMI ($r=-0.06$, NS). The DISC Total score was significantly correlated with the SS Total score ($r=0.26$, $p<0.05$) and Discrimination score ($r=0.38$, $p<0.001$), with the latter reaching threshold.

3.2.3.3. Divergent validity. Divergent validity was assessed by considering the association between gender and DISC Total score. No significant relationship was found ($t=-1.89$, $p=0.063$).

3.2.4. Precision

3.2.4.1. Corrected item-total correlations. The total corrected item-total correlation (r_s) for the Unfair Treatment subscale was 0.38. Correlations between items and the subscale corrected item totals ranged from $r_s=0.55$ for 'treated unfairly making or keeping friends' to $r_s=0.24$ for 'treated unfairly in marriage or divorce', which had a correlation lower than the predefined threshold of $r_s=0.3$. The total corrected item-total correlation for the Stopping Self subscale was $r_s=0.66$. Correlations between items and the subscale total ranged from $r_s=0.50$ for 'stopped yourself from having a close personal relationship' to $r_s=0.66$ for 'stopped yourself from applying for education or training courses'. No items had a correlation lower than the predefined threshold of $r_s=0.3$. The total corrected item-total correlation for the Overcoming Stigma subscale was $r_s=0.20$. The correlation between the item and subscale total was $r_s=0.20$ for each item. No items met the predefined threshold. The total corrected item-total correlation for the positive treatment subscale was $r_s=0.67$. Correlations between items and the subscale total ranged from $r_s=0.59$ for 'treated more positively by mental health staff' to $r_s=0.68$ for 'treated more positively by your family'. No items had a correlation lower than the predefined threshold.

3.2.4.2. Correlation with own subscale and others. All items, with the exception of 'use your personal skills or abilities in coping with stigma', correlated more highly with their own subscale than any of the other three subscales. For the Unfair Treatment subscale, correlations of items with the subscale ranged from $r=0.25$,

$p<0.05$ for 'treated unfairly in marriage or divorce' to $r=0.59$, $p<0.01$ for 'treated unfairly making or keeping friends'. For the Stopping Self subscale, all items correlated more highly with this subscale rather than any of the others. Correlations of items with the subscale ranged from $r=0.36$, $p<0.01$ for 'stopped yourself from applying for education or training courses' to $r=0.57$, $p<0.01$ for 'stopped yourself from having a close personal relationship'. For the Overcoming Stigma subscale, 'made friends with people who do not use mental health services' was correlated with the subscale total at $r=0.33$, $p<0.01$ while 'use your personal skills or abilities in coping with stigma' was correlated at $r=0.31$, $p<0.01$. For the Positive Treatment subscale, correlations of items with the subscale ranged from $r=0.29$, $p<0.01$ for 'treated more positively in employment' to $r=0.56$, $p<0.01$ for 'treated more positively by mental health staff'.

3.2.4.3. Acceptability. Several of the items had a high endorsement of the non-applicable response, e.g., 'treated unfairly in your role as a parent to your children' at 67.1% and 'treated unfairly in marriage or divorce' at 58.8%. The MEF criterion was not violated as no response categories contained $>=80\%$ of responses. There were no items which violated the AEF when considering the adjacent categories of 'not at all' and 'a little'. There were three items which violated the AEF when considering the adjacent categories of 'a little' and 'moderately'; 'treated unfairly in your role as a parent to your children'; 'treated unfairly in marriage or divorce'; 'treated unfairly in starting a family or having children'. Of these items, the non-applicable response option was frequently endoresed at 67.1%, 58.8% and 42.9% respectively. One item violated the AEF when considering the adjacent categories of 'moderately' and 'a lot'; 'treated unfairly in religious practices'. This also had a large endorsement of the 'not at all' category at 60.7% of responses.

3.2.5. Feasibility

DISC completion time ranged from 5.0 min to 85.3 min. When two outlying cases (82.9 min and 85.3 min) were removed, the mean completion time was 29.1 min (S.D.=12.8). The median completion time was 28.4 min.

4. Discussion

This study conducted further developmental work and evaluated the psychometric properties of the DISC. The DISC was designed to address the need for a psychometrically validated measure which considers the scope and content of experienced discrimination. Experienced discrimination was conceptualised as reported instances of unfair treatment or discrimination due to having a mental health problem. This work led to a recommendation for the 22-item Unfair Treatment subscale to be used as a stand-alone measure of experienced discrimination. This subscale was psychometrically robust, fully meeting reliability, validity, precision and feasibility criteria. The only area in which it did not meet the threshold was the AEF in which four items violated the criterion. This violation is not necessarily a cause for concern, however, as these items detail areas of life which were not applicable to all participants including parenting, marriage and religion. The high proportion of non-applicable responses for these items was a contributing factor to the violation of the AEF. These items were considered sufficiently important to include despite their lack of universal applicability, based on participant responses in both the earlier DISC studies and cognitive debriefing and focus groups conducted as part of this study (Thornicroft et al., 2009; Rose et al., 2011). Therefore, there is a conceptual justification for their inclusion despite this psychometric violation.

The reliability of other DISC subscales and the total score was satisfactorily established, including internal consistency, test–re-test reliability and inter-rater reliability. The criterion for feasibility was also established. Elements of validity, precision, and acceptability were established, and areas leading to the further improvement of these properties were also highlighted. The Overcoming Stigma and Positive Treatment subscales did not correlate at a moderately level with the Stopping Self and Unfair Treatment subscales, suggesting that they may be measuring a different construct to that assessed in the other subscales. The lack of convergent validity between the Stopping Self subscale and the Disclosure subscale of the SS and the Social Withdrawal subscale of the ISMI suggests that further work is needed on this subscale if it is to represent these constructs adequately. Further work on these three subscales is recommended to improve their validity and precision. However, they are sufficiently reliable, acceptable and feasible for current use. Further work is underway to develop a new measure of anticipated discrimination—the Questionnaire on Anticipated Discrimination (QUAD). This measure will be designed with the aim of using it to replace the Stopping Self, Overcoming Stigma and Positive Subscales of the DISC. These subscales will be used as the initial basis of the new scale. Therefore, it is recommended to drop these elements from the DISC, and, use it as a 22-item measure of experienced stigma, as currently represented by the Unfair Treatment subscale. Further work will also be necessary to evaluate the psychometric properties of the DISC in additional clinical populations or in cultural groups other than those included in this article.

4.1. Strengths and limitations

The psychometric evaluation of the DISC has been strengthened by building on the assessment of elements of acceptability and feasibility, as established in the further development work. The 22-item scale contributes evidence to support the evaluation of outcomes as part of social inclusion interventions or anti-stigma campaigns, as indicated by the Medical Research Council's guidance on developing and evaluating complex interventions (Craig et al., 2008).

The qualitative element of the DISC is a strength as it offers researchers a structured method to consider the types of experiences which are reported as discriminating, as well as the frequency and magnitude of these experiences which are represented in the quantitative elements of the scale. In previous work using the DISC, it was found that the types of experiences reported as 'positive discrimination' were very different in magnitude to those reported as 'negative discrimination'. Many of the negative examples were 'strong'; concerning experiences such as rejection, teasing, humiliation, abuse. The 'positive' examples, on the other hand, never contained such strong connotations with examples generally referring to other people treating the participant well—they were treated 'nicely', with kindness, with understanding (Rose et al., 2011). This is interesting in the context of studies which suggest that reports of experienced discrimination are limited by the social costs of making attributions to discrimination (e.g., Kaiser and Miller, 2001). Attributing a lack of success in areas of life to stigma may lead to negative outcomes for the individual such as a lack of perceived control over one's environment or a lowering of social self-esteem. This suggests that the lesser focus on these areas may mirror the individual's tendency to perceive stigma but remain cautious in defining experiences of less favourable treatment as discrimination or in applying these perceptions to themselves. It is useful to consider that this phenomenon may lead to an under-reporting of negative experienced discrimination. Although, in our current study, all but one participant endorsed at least one aspect in life in which they had experienced stigma (98.8%), mean levels were generally in the minimal category with an average score for magnitude

of stigma of 0.9 out of a maximum score of 4. This suggests that this underreporting phenomenon may require consideration when interpreting experienced discrimination scores.

This study is limited as data collection was not powered to perform factor analysis. This decision was taken as this psychometric requirement can be fulfilled by performing secondary analysis on an existing dataset. This additional dataset was, however, not sufficient to conduct the analyses described here as no additional measures were administered, multiple ratings were not performed and data collection was at one time point only. Secondary analysis of this dataset will allow the factor structure of the DISC to be established, complementing the analyses reported in this article. Secondary analysis of the additional dataset will also allow an item response theory approach to be taken, particularly the investigation of Differential Item Functioning. ANOVA will be conducted for each item to examine uniform and non-uniform Differential Item Functioning across each level of the 'person factor' (employment status) and across different levels of the DISC constructs (class intervals). Again, the sample size of this current study is not sufficient to perform these analyses.

The study is also limited as all data were collected from one NHS trust in South London. This study was designed to establish the psychometric properties of the DISC, and such a sample was adequate for this purpose. The sample is diverse and contains individuals from a range of socio-economic and ethnic backgrounds, as well as diagnoses. However, all were urban dwellers. Further work is necessary to establish the degree to which the experiences reported by this sample are replicated in other areas of the UK, and internationally.

4.2. Conclusions

In conclusion, the 22-item DISC is a reliable, valid, precise, acceptable, and feasible measure for use in assessing experienced discrimination. The use of this scale is recommended as an evaluation tool. The Stopping Self, Overcoming Stigma and Positive Treatment subscales require further modification before they can be recommended as meeting all psychometric requirements; therefore, they are not included in the recommended version of the DISC. Further work is underway to develop these subscales into a stand-alone measure.

Acknowledgements

Data collection assistance was given by Amy MacDonald, Gaby Illingsworth, Zoe Given-Wilson, Eleanor Parker, Manuela Jarrett, Tanya Graham, Georgia Black and Eleanor Lewis-Holmes. Dr Morven Leese provided statistical advice. This article presents independent research commissioned by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research scheme (RP-PG-0606–1053). The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. GT is also funded through a NIHR Specialist Mental Health Biomedical Research Centre at the Institute of Psychiatry, King's College London and the South London and Maudsley NHS Foundation Trust.

References

Angermeyer, M.C., Dietrich, S., 2006. Public beliefs about and attitudes towards people with mental illness: a review of population studies. Acta Psychiatrica Scandinavica 113, 163–179.

Brohan, E., Elgie, R., Sartorius, N., Thornicroft, G., 2010a. Self-stigma, empowerment and perceived discrimination among people with schizophrenia in 14 European countries: the GAMIAN-Europe study. Schizophrenia Research 122, 232–238.

Brohan, E., Gauci, D., Sartorius, N., Thornicroft, G., 2011. Self-stigma, empowerment and perceived discrimination among people with bipolar disorder or depression in 13 European countries: the GAMIAN-Europe study. Journal of Affective Disorders 129, 56–63.

Brohan, E., Slade, M., Clement, S., Thornicroft, G., 2010b. Experiences of mental illness stigma, prejudice and discrimination: a review of measures. BMC Health Services Research 10, 80.

Craig, P., Dieppe, P., MacIntyre, S., Mitchie, S., Nazareth, I., Petticrew, M., 2008. Developing and evaluating complex interventions: the new Medical Research Council guidance. British Medical Journal 337, 979–983.

Crocker, J., Major, B., Steele, C., 1998. Social stigma. In: Gilbert, D.T., Fiske, S.T., Lindzey, G. (Eds.), Handbook of Social Psychology. McGraw-Hill, Boston, pp. 504–533.

Cronbach, L.J., 1951. Coefficient alpha and the internal structure of tests. Psychometrika 16, 297–334.

Endicott, J., Spitzer, R., Fleiss, J., Cohen, J., 1976. The global assessment scale. Archives of General Psychiatry 33, 765–771.

Fitzpatrick, R., Davey, C., Buxton, M.J., Jones, D.R., 1998. Evaluating patient based outcome measures for use in clinical trials. Health Technology Assessment 2 (14).

Flesch, R., 1974. The Art of Readable Writing. Harper and Row, New York.

Joffe, H., Yardley, L., 2004. Content and thematic analysis. In: Marks, D.F., Yardley, L. (Eds.), Research Methods for Clinical and Health Psychology. SAGE publications, London, pp. 56–68.

Jones, S.H., Thornicroft, G., Coffey, M., Dunn, G., 1995. A brief mental health outcome scale-reliability and validity of the Global Assessment of Functioning (GAF). British Journal of Psychiatry 166, 654–659.

Kaiser, C.R., Miller, C.T., 2001. Stop complaining! The social costs of making attributions to discrimination. Personality and Social Psychology Bulletin *27* (2), 254–263.

Kalk, N.J., Pothier, D.D., 2008. Patient information on schizophrenia on the internet. The Psychiatrist 32, 409–411.

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. British Journal of Psychiatry 190, 248–254.

Lamping, D.L., Schroter, S., Marquis, P., Marrel, A., Duprat-Lomon, I., Sagnier, P.P., 2002. The Community-Acquired Pneumonia Symptom Questionnaire*: a new, patient-based outcome measure to evaluate symptoms in patients with community-acquired pneumonia. Chest 122, 920–929.

Leucht, S., Kane, J.M., Kissling, W., Hamann, J., Etschel, E., Engel, R., 2005. Clinical implications of Brief Psychiatric Rating Scale scores. British Journal of Psychiatry 187, 366–371.

Lin, L.I.K., 1989. A concordance correlation coefficient to evaluate reproducibility. Biometrics 45, 255–268.

Link, B.G., Struening, E.L., Rahav, M., Phelan, J.C., Nuttbrock, L., 1997. On stigma and its consequences: evidence from a longitudinal study of men with dual diagnosis of mental illness and substance abuse. Journal of Health and Social Behavior 38, 177–190.

Lukoff, D., Neuchtereiein, K.H., Ventura, J., 1986. Appendix A. Manual for expanded Prief Psychiatric Rating Scale (BPRS). Schizophrenia Bulletin 12, 594–602.

Lysaker, P.H., Roe, D., Yanos, P.T., 2007. Toward understanding the insight paradox: Internalized stigma moderates the association between insight and social functioning, hope, and self-esteem among people with schizophrenia spectrum disorders. Schizophrenia Bulletin 33, 192–199.

Ojanen, V., Gogates, G., 2006. A briefing on cognitive debriefing. Good Clinical Practice Journal 13, 25–29.

Overall, J.E., Gorham, D.R., 1962. The Brief Psychiatric Rating Scale. Psychological Reports 10, 799–812.

Ritsher, J.B., Otilingam, P.G., Grajales, M., 2003. Internalized stigma of mental illness: psychometric properties of a new measure. Psychiatry Research 121, 31–49.

Rose, D., Willis, R., Brohan, E., Sartorius, N., Thornicroft, G., Villares, C., Wahlbeck, K., 2011. the INDIGO Study Group, 2011. Reported stigma and discrimination by people with a diagnosis of schizophrenia. Epidemiology and Psychiatric Sciences 20, 193–204.

SPSS Inc., 2006. SPSS for Windows, Release 15.0. SPSS, Inc., Chicago.

StataCorp, 2005. Stata Statistical Software: Release 9.2. StataCorp LP, College Station, TX.

Terwee, C.B., Bot, S.D.M., de Boer, M.R., van der Windt, D.A.W.M., Knol, D.L., Dekker, J., Bouter, L.M., de Vet, H.C.W., 2007. Quality criteria were proposed for measurement properties of health status questionnaires. Journal of Clinical Epidemiology 60, 34–42.

The Whoqol Group, 1998. The World Health Organization quality of life assessment (WHOQOL): development and general psychometric properties. Social Science & Medicine 46, 1569–1585.

Thornicroft, G., Brohan, E., Rose, D., Sartorius, N., Leese, M., 2009. The INDIGO Study Group, 2009. Global pattern of experienced and anticipated discrimination against people with schizophrenia: a cross-sectional survey. The Lancet 373, 408–415.

Thornicroft, G., Rose, D., Kassam, A., Sartorius, N., 2007. Stigma: ignorance, prejudice or discrimination? British Journal of Psychiatry 190, 192–193.

Wahl, O.F., 1999. Mental Health consumers experience of stigma. Schizophrenia Bulletin 25, 467–478.

Walter, S.D., Eliasziw, M., Douglas, J.A., 1998. Sample size and optimal designs for reliability studies. Statistics in Medicine 17, 101–110.

Yanos, P.T., Roe, D., Markus, K., Lysaker, P.H., 2008. Pathways between internalised stigma and outcomes related to recovery in schizophrenia spectrum disorders. Psychiatric Services 59, 1437–1442.