

# The quality of mental health literacy measurement tools evaluating the stigma of mental illness: a systematic review

Y. Wei<sup>1\*</sup>, P. McGrath<sup>2</sup>, J. Hayden<sup>3</sup> and S. Kutcher<sup>4</sup>

<sup>1</sup> Faculty of Graduate Studies, Interdisciplinary PhD, Dalhousie University, Halifax, Nova Scotia, Canada

<sup>2</sup> IWK Health Centre, Halifax, Nova Scotia, Canada

<sup>3</sup> Faculty of Medicine, Dalhousie University, Halifax, Nova Scotia, Canada

<sup>4</sup> Department of Psychiatry, Faculty of Medicine, Dalhousie University, Halifax, Nova Scotia, Canada

**Aims.** Stigma of mental illness is a significant barrier to receiving mental health care. However, measurement tools evaluating stigma of mental illness have not been systematically assessed for their quality. We conducted a systematic review to critically appraise the methodological quality of studies assessing psychometrics of stigma measurement tools and determined the level of evidence of overall quality of psychometric properties of included tools.

**Methods.** We searched PubMed, PsycINFO, EMBASE, CINAHL, the Cochrane Library and ERIC databases for eligible studies. We conducted risk-of-bias analysis with the Consensus-based Standards for the Selection of Health Measurement Instruments checklist, rating studies as excellent, good, fair or poor. We further rated the level of evidence of the overall quality of psychometric properties, combining the study quality and quality of each psychometric property, as: strong, moderate, limited, conflicting or unknown.

**Results.** We identified 117 studies evaluating psychometric properties of 101 tools. The quality of specific studies varied, with ratings of: excellent ( $n=5$ ); good (mostly on internal consistency ( $n=67$ )); fair (mostly on structural validity,  $n=89$  and construct validity,  $n=85$ ); and poor (mostly on internal consistency,  $n=36$ ). The overall quality of psychometric properties also varied from: strong (mostly content validity,  $n=3$ ), moderate (mostly internal consistency,  $n=55$ ), limited (mostly structural validity,  $n=55$  and construct validity,  $n=46$ ), conflicting (mostly test-retest reliability,  $n=9$ ) and unknown (mostly internal consistency,  $n=36$ ).

**Conclusions.** We identified 12 tools demonstrating limited evidence or above for (+, ++, +++) all their properties, 69 tools reaching these levels of evidence for some of their properties, and 20 tools that did not meet the minimum level of evidence for all of their properties. We note that further research on stigma tool development is needed to ensure appropriate application.

Received 1 November 2016; Accepted 26 March 2017; First published online 2 May 2017

**Key words:** mental illness stigma, psychometrics, systematic reviews, validation study.

## Introduction

Approximately 50–85% of people with severe mental disorders receive no treatment (Patel *et al.* 2007; World Health Organization, 2011). People with mental illness have difficulty accessing mental health care due to many factors, amongst which stigma against mental illness is one significant barrier, according to a recent systematic review on variables influencing mental health help-seeking (Gulliver *et al.* 2010).

Stigma of mental illness is ‘a trait that is deeply discrediting that reduces the bearer from a whole to a

tainted, discounted one’ (Goffman, 1963). Several conceptual frameworks have been created, including labelling theory (Goffman, 1963; Link *et al.* 1987), social attribution theory (Corrigan *et al.* 2003), cognitive behavioural modelling (Thornicroft, 2006) and social stigma modelling (Jones *et al.* 1984), to both help understand and evaluate stigma related to mental illness, and guide stigma reduction interventions. As a result, the dimensions of the stigma of mental illness vary from one theory to another, and so do the stigma measurement tools created under different theories. More recently, the mental health literacy framework (Kutcher *et al.* 2015a, b, 2016) considers stigma reduction as one of its core constructs and stresses how stigma reduction and the improvement of mental health knowledge may enhance help-seeking behaviours. Research, such as randomised controlled trials

\* Address for correspondence: Y. Wei, Faculty of Graduate Studies, Interdisciplinary PhD, Dalhousie University, Halifax, Nova Scotia, Canada.

(Email: yifeng.wei@iwk.nshealth.ca)

and longitudinal cohort studies (McLuckie *et al.* 2014; Kutcher *et al.* 2015a, b; Milin *et al.* 2016; Thornicroft *et al.* 2016) have demonstrated the effectiveness of interventions designed based on this approach.

Under these frameworks, a plethora of measurement tools have been developed to evaluate the stigma of mental illness from different lenses. This includes the evaluation of public stigma/personal stigma, people's own attitudes towards people with mental illness; perceived stigma that people perceive as held by others towards people with mental illness; self-stigma that people with mental illness hold against themselves; and experienced stigma that people with mental illness have encountered at the individual, community and society levels (Batterham *et al.* 2013). A recent scoping review (Wei *et al.* 2015), a systematic approach to map the literature in an area of interest and to accumulate and synthesise evidence available, identified 65 stigma measures and a narrative review (Brohan *et al.* 2010) identified another 14, and categorised them according to different theoretical models. Another narrative review discussed more than 100 stigma measures informed by labelling theory specifically (Link *et al.* 2004). One narrative review (Boyd *et al.* 2014) discussed 47 versions of one tool, Internalized Stigma of Mental Illness, and summarised related reliability and validity. However, despite the abundance of stigma measurement tools, and stigma impact research using them, there has been little, if any, research identified to investigate the quality of currently available stigma measurement tools. Furthermore, this has been no research identified to aggregate, analyse and compare stigma measurement tools developed under different stigma theoretical frameworks.

We conducted a systematic review to critically analyse the methodological quality of studies on psychometrics of available stigma tools and further to determine the level of evidence of the overall quality of their psychometrics across studies. Based on our analysis we then make recommendations for further stigma research and the application or ongoing development of these tools.

## Methodology

This review followed the protocol recommended by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (Moher *et al.* 2009) to report its findings. We conducted risk of bias analysis with the adapted Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) checklist (Terwee *et al.* 2012); assessed the quality of each individual psychometric property,

using criteria developed by the COSMIN group (Terwee *et al.* 2007); and then rated the level of evidence of overall quality. COSMIN checklist is a consensus-based checklist used to evaluate the methodological quality of studies on the measurement properties of health status instruments (Terwee *et al.* 2012).

## Search strategy

We searched the databases of PubMed, PsycINFO, EMBASE, CINAHL, the Cochrane Library and ERIC for relevant studies without limit on publication dates. The search period was between January and June 2015 and updated the search between April and May 2016, assisted by a local university health librarian. To ensure our search covered all dimensions of stigma as framed within the mental health literacy approach, regardless of theoretical foundations they were affiliated with, our search strategy covered all three outcomes of mental health literacy (knowledge, stigma and help-seeking) and we did not exclude studies that self-identified as focused on knowledge or help-seeking outcomes until the last stage of data extraction because some mental health literacy measures include all three components. We applied the search strategy from the scoping review (Wei *et al.* 2015) that contained four sets of key words and phrases regarding general mental health and mental disorders, three outcomes of mental health literacy, assessment tools and study designs. Appendix 1 provides details of all search words and phrases applied searching PubMed.

Two team members independently searched the citations identified from database searches for relevant studies. Both members followed the same procedures to assess potential relevance of studies: reviewing titles in general (stage 1), reviewing titles and scanning abstracts (stage 2), briefly scanning full papers (stage 3) and reading full papers for data extraction (stage 4). Following these stages, we checked the reference list of each included study for additional studies and further searched narrative reviews on stigma measurement tools for additional studies (Link *et al.* 2004; Brohan *et al.* 2010; Boyd *et al.* 2014). The two reviewers discussed their identified studies and reached consensus on the final inclusion of studies. Three mental health professionals and/or research methodologist were available to solve any discrepancies on the final decisions for included studies.

## Selection criteria

We included any type of quantitative studies assessing and reporting any psychometrics (reliability, validity

and responsiveness) of a stigma measurement tool. According to the literature review, we defined that a stigma measurement tool evaluates: perceived stigma, experienced stigma, emotional responses to mental illness or self-stigma of mental illness. Our search focused on tools addressing stigma of mental illness in general or stigma against common specific mental illnesses: anxiety disorder, depression, attention deficit hyperactivity disorder (ADHD) and schizophrenia. An eligible study had to report not only the psychometrics of the tool, but also the statistical analysis of these psychometrics. We searched databases for studies published in English and did not limit the date of publication, or study participant age.

We excluded studies that only provided psychometrics of the tool applied, but did not report the statistical analysis of these psychometrics. For example, many studies evaluating anti-stigma interventions reported the internal consistency of the tool applied but did not describe the statistical analysis related to it and therefore were excluded from our review. We did not include studies addressing stigma related to substance use and addictions as they cover a wide range of domains that need independent evaluation.

#### **Data extraction**

We followed the COSMIN checklist manual (Terwee *et al.* 2012) and created a data extraction form *a priori* to document basic information of each included study, such as author information, the tool content, response option of the tool, population, study location and study sample size. We further documented information about measurement properties as: (1) reliability (internal consistency, reliability (test-retest and intra-rater reliability) and measurement errors); (2) validity (content validity, structural validity (factor analysis), hypothesis testing (construct validity), cross-cultural validity and criterion validity); and (3) responsiveness (sensitivity to change).

We considered adapted tools (adding/reducing items or changing original items) as separate tools. However, if a tool was created in one study but in another was assessed for its factors and the number of final items was adjusted from the original tool due to the factor analysis, we considered them as the same tool as this is part of the usual ongoing process of finalising scales.

#### **Methodological quality of included studies (risk of bias)**

We rated the quality of a study for a particular measurement property as: 'excellent', 'good', 'fair' or 'poor'. As a study may assess more than one

measurement property, it may have multiple levels of quality for different measurement properties it assesses. The COSMIN checklist (Terwee *et al.* 2012) created 7–18 criteria items to assess the methodological study quality for each measurement property, rated as 'excellent', 'good', 'fair' or 'poor' under each item, respectively. The final ranking of the study quality for each property takes the lowest criteria ranking. For example, the COSMIN checklist contains seven criteria items to assess the study quality assessing structural validity, and if under each item the study has different ranking ranging from 'poor' to 'good', the final ranking for this study would be 'poor' for structural validity.

#### **Quality of measurement property and level of evidence of overall quality**

In addition, the COSMIN group developed quality criteria for each psychometric property (except for cross-cultural validity) (Terwee *et al.* 2007). Each property must reach a quality threshold to receive a positive rating (+), otherwise a negative rating (–) or indeterminate rating due to the lack of data (?), or conflicting rating (+/–) if the findings are contradictory (Appendix 2). Based on both the methodological study quality and the quality of each psychometric property, we determined the level of evidence of overall quality of a psychometric property. The ratings were determined by adapting and applying criteria from a systematic review on measures of continuity of care (Uijen *et al.* 2012) and the Cochrane Back and Neck Group's recommendations on the overall level of evidence of each assessed outcome (Furlan *et al.* 2015) (Appendix 3). As a result, the levels of evidence are: strong (S) (+++ or ---), moderate (M) (++ or --), limited (L) (+ or –), conflicting (C) (+/–), or unknown (U) (x). We considered measurement properties with positive strong evidence (+++) as 'ideal', moderate positive evidence (++) as 'preferred', and limited positive evidence (+) as 'minimum acceptable'.

We defined the level of evidence as unknown (U(x)) if: (1) a property is assessed in one study only and the study quality is 'poor', or the psychometric property is indeterminate (?); (2) a property is assessed in two studies, and the study quality is poor or property is indeterminate (?) in both studies; (3) a property is assessed in more than two studies, and the study quality is poor or property is indeterminate (?) in  $\geq$  half of the studies.

If a property is assessed in two studies and study quality is  $\geq$  'fair', and the quality of the measurement property is positive (+) in both studies, we used the 'worst score' approach for the level of evidence, otherwise we determined the level of evidence as conflicting

(C(+/-)). If a property is assessed in more than two studies and we found fair, good or excellent study quality in more than half of the studies, we considered the level of evidence as strong, moderate or limited, using the 'worst score account' approach. For example, if a measurement property is rated as (+) or (-) consistently in studies with the mixed study quality of excellent, good and fair, the final rating is limited level of evidence (L(+) or L(-)). For the rest of the cases, the level of evidence is conflicting (C (+/-)).

## Results

### Study selection and characteristics

Figure 1 presents the flow chart of study selection process. The data were first imported into Reference 2.0 database management software (RefWorks-COS PL, ProQuest, 2001) and duplicates were removed. We then screened 21 089 studies, and excluded studies that were not the topic of interest (e.g., studies addressing HIV/AIDS stigma, CBT, resilience, social and emotional learning, mental disorders that were not the topic of interest of this review) through four screening stages. As a result, we identified 117 studies reporting and analysing psychometric properties of 101 stigma measurement tools (Table 1). We classified tools according to what they measured (Table 1):

perceived stigma against mental illness or the mentally ill; perceived stigma against mental health care (e.g., treatment, help-seeking, mental health institutions or psychiatry as a profession); emotional responses to mental illness; experienced stigma by people with mental illness or their relatives/caregivers; self-stigma by people with mental illness. We did not categorise tools under a specific stigma theory because most were developed with combined components from various theories or based on interviews with target population.

Ninety-one out of 101 tools applied Likert-scale response format asking participants to rate the level of agreement on items addressing stigma (Table 1). The other 10 tools applied formats such as multiple choices (e.g., yes/no/do not know); responses on a 100 mm visual analogue scale; error-choice response; open-ended questions; or prevalence and frequency of stigma experience.

Study participants were mostly people with mental illness ( $n = 36$ ) and their relatives and caregivers ( $n = 6$ ), followed by community members/general public ( $n = 20$ ), health care providers and staff ( $n = 20$ ), college students ( $n = 15$ ), secondary school students ( $n = 8$ ); and people from other professions such as educators ( $n = 2$ ), police ( $n = 1$ ), athletes ( $n = 1$ ), employers ( $n = 1$ ) and military personnel and veterans ( $n = 1$ ). Some studies used multiple groups of participants mentioned

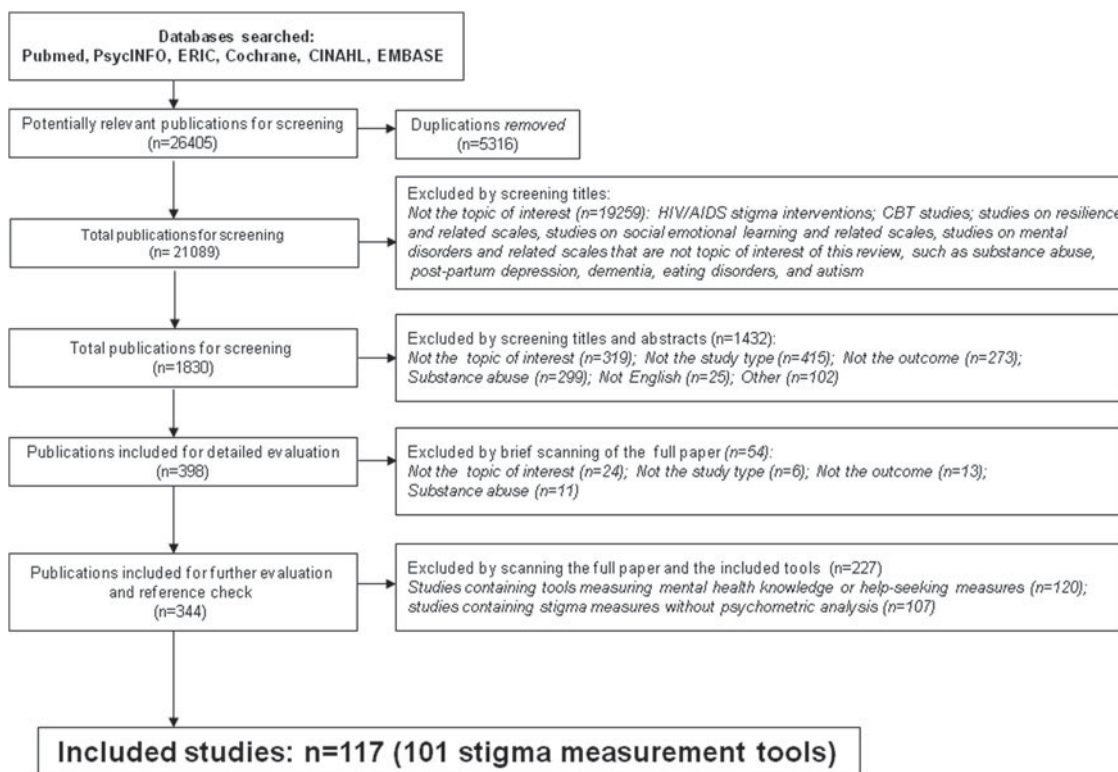


Fig. 1. Flow chart of search results.

**Table 1.** Study characteristics

Measurement tools	Response options	Population	Age	Sample	Country	Content
1. Self-reported prejudiced attitudes (SRPA) (Andersson <i>et al.</i> 2010)	9 4-point scaled items	Upper secondary schools	16–19	4046	Norway	A
2. Personal attributes (PR) (Angermeyer, Matschinger, 2003)	8 5-point scaled items	Community members	≥18	5025	German	A
3. Labelling (Angermeyer & Matschinger, 2003)	Open-ended questions for a vignette	Community members	≥18	5025	German	A
4. Peer Mental Health Stigmatization scale (PMHSS) (McKeague <i>et al.</i> 2015)	24 5-point scaled items	Children and adolescents in schools	M = 12.99 ± 1.6	562	Ireland	A
5. Knowledge Test of Mental Illness (KT) (Michaels & Corrigan, 2013)	14 items with numerical or true/false responses	College students, community members, mental health providers and consumers	M = 21.6 ± 3.2; 33.1 ± 7.4; 45.5 ± 11.4; 45.4 ± 11.2	35; 203; 133; 74	US	A
6. Day's Mental Illness scale (DMIS) (Day <i>et al.</i> 2007)	28 7-point scaled items	College students, community members and people with mental illness	M = 24.84; 18.60; 45	341; 42; 20	US	A
7. EMIC [45] (Chowdhury <i>et al.</i> 2000)	13 4-point scaled items	Laypersons and health care providers	?	38	India	A
8. Employer Attitudes towards Mental Illness questionnaire (EAQ) (Diksa & Rogers, 1996)	38 5-point scaled items	Employers	?	373	US	A
9. Attitudes of Nursing Staff towards Co-Workers Returning from Psychiatric and Physical Illnesses (ANCW) (Glozier <i>et al.</i> 2006)	12 4-point scaled items	Nursing staff	M = 35.6–38.6	117	UK	A
10. Depression Stigma scale (DSS)/PPSS (Griffiths <i>et al.</i> 2004)	18 5-point scaled items	Community members	M = 36.4 ± 9.4	525	Australia	A
DSS (Griffiths <i>et al.</i> 2008)	18 5-point scaled items	Community members	Median = 45–49; M = 35.9 ± 9.2; 35.3 ± 8.76	1001; 5572; 487	Australia	A
11. DSS revised (Gulliver <i>et al.</i> 2012)	9 5-point scaled items	Elite athletes	M = 25.5	59	Australia	A
12. Generalized Anxiety Stigma scale (GASS) (Griffiths <i>et al.</i> 2011)	20 5-point scaled items	General public	M = 46.6 ± 13.25	617; 212	Australia	A

Continued

Table 1. Continued

Measurement tools	Response options	Population	Age	Sample	Country	Content
13. GASS (Gulliver <i>et al.</i> 2012)	10 5-point scaled items	Elite athletes	M = 25.5	59	Australia	A
14. Test of Knowledge About ADHD (KADD) (Hepperlen <i>et al.</i> 2002)	22 error-choice items	Elementary school teachers	M = 39.43	103	US	A
15. Beliefs toward Mental Illness (BMI) (Hirai & Clum, 2000)	24 6-point scaled items	College students	M = 25.3 ± 5.1	216	US	A
16. Opening Minds Scale for Health Care Providers (OMS-HC) (Kassam <i>et al.</i> 2012)	20 5-point scaled items	Health care providers	≥18	787	Canada	A
OMS-HC [111] (Modgill <i>et al.</i> 2014)	20 5-point scaled items	Health care providers	18–65	1523	Canada	A
17. ADHD Stigma Questionnaire (ASQ) (Kellison <i>et al.</i> 2010)	26 4-point scaled items	Adolescents at risk of ADHD	M = 15.6 ± 1.8	301	US	A
ASQ (Bell <i>et al.</i> 2011)	26 4-point scaled items	Teachers	M = 42.32 ± 12.61	268	US	A
18. ADHD Stigma (Fuermaier <i>et al.</i> 2012)	37 6-point scaled items	College students, community members, teachers and physicians	M = 31.3 ± 14.8; M = 50.6–52.3	1033; 228	Netherlands	A
19. SS Chinese version (Ho <i>et al.</i> 2015)	14 5-point scaled items	People with mental illness	M = 51.2 ± 11.34	114	China	A
20. Self-Esteem and Stigma Questionnaire (SE/SQ) (Hayward <i>et al.</i> 2002)	8 6-point scaled items	Patients with depression	M = 43 ± 11	186	UK	A
21. Devaluation-Discrimination tool (DD) (Link, 1987)	12 4-point scaled items	Community members and people with mental illness	M = 32.71–40.29	593	US	A
Perceived DD (Bjorkman <i>et al.</i> 2007)	12 4-point scaled items	Patients with mental illness	M = 46	40	Sweden	A
Perceived DD (depression) (Interian <i>et al.</i> 2010)	12 4-point scaled items	Latino primary care patients	≥18	200	US	A
22. Depression is a Matter of Will (Aromaa <i>et al.</i> 2010)	16 4-point scaled items	Community members	M = 46.9 ± 17.3	5520	Finland	A
23. Public stigma/PDD (Moses, 2009)	14 4-point scales items	Youth with mental illness	12–18	60	US	A
24. Perceived dangerousness (PD) (Link <i>et al.</i> 1987)	8 6-point scaled items	Community members	M = 47.6	152	US	A
25. PD/Link Stigma scale (Bagley & King, 2005)	11 5-point scaled items	Patients with mental illness	M = 42.2	83	UK	A
26. Dangerousness scale (Penn <i>et al.</i> 1994)	8 7-point scaled items	College students	?	329	US	A
27. Social distance (SD) (Link <i>et al.</i> 1989)	7 4-point scaled items	Community members	M = 47.6	152	US	A
SD (Penn <i>et al.</i> 1994)	7 4-point scaled items	College students	?	329	US	A

28. SD revised (Interian <i>et al.</i> 2010)	Six multiple choice items	Latino primary care patients	≥18	200	US	A
29. Reported and Intended Behaviour scale (RIBS) (Evans-Lacko <i>et al.</i> 2013)	Eight items	General public	M = 46 ± 18.6	6954	UK	A
RIBS (Evans-Lacko <i>et al.</i> 2012)	Four items on an ordinal scale and 4 items on multiple choices	General public	M = 38.1 ± 13.4; 36.9 ± 14.1	403; 83	UK	A
RIBS (Evans-Lacko <i>et al.</i> 2011)	Four items on an ordinal scale and 4 items on multiple choices	General public	25–45	92; 37; 403	UK	A
RIBS (Friedrich <i>et al.</i> 2013)	Four items on an ordinal scale and 4 items on multiple choices	Medical students	M = 23.5	1452	UK	A
RIBS Japanese version (Yamaguchi <i>et al.</i> 2014)	Four items on an ordinal scale and 4 items on multiple choices	Undergraduate and graduate students	M = 22.61 ± 2.47	224	Japan	A
30. Social Contact scale (SD version) (Jackson & Heatherington, 2006)	8 4-point scaled items	Secondary school students	M = 13.3 ± 1.26	1223	Jamaica	A
31. Social Supports Acceptance scale (SSAS) (Mansouri & Dowell, 1989)	? 4-point scaled items	People with mental illness	18–70	70	US	A
32. Attitudes to Mental Illness Questionnaire (AMID) (Luty <i>et al.</i> 2006)	5 5-point scaled items	General public	M = 46.3 ± 15.7	1079	UK	A
33. Attitudes to Severe Mental Illness (ASMI) (Madianos <i>et al.</i> 2012)	35 6-point scaled items	General public	M = 41.5 ± 10.61; 43.71 ± 11.18	2039	Greece	A
34. Attitudes Toward Social Competence and Integration of People with Mental Illness (ASCI) (Minnebo & Acker, 2004)	8 5-point scaled items	Belgian high school students	M = 16.8 ± 1.6	207	Belgium	A
35. Client Attitude Questionnaire (CAQ) (Morrison & Becker, 1975)	20 true/false/do not know items	Psychiatric professionals	?	13	US	A
36. Beliefs and attitudes toward people diagnosed with psychosis (BAP) (Serra <i>et al.</i> 2013)	6 5-point scaled items	High school students	M = 17.3 ± 1.4; 17.3 ± 1.3	1023	Italy	A
37. Devaluation of consumer family scale (DSFS) (Struening <i>et al.</i> 2001)	15 4-point scaled items	Caregivers of people with mental illness	M = 50 ± 14.3	461	US	A

Continued

Table 1. Continued

Measurement tools	Response options	Population	Age	Sample	Country	Content
Stigma-Devaluation scale (Dalky, 2012)	15 4-point scaled items	Family members of people with mental illness	M = 44.5 ± 11.7	164	Jordan	A
38. Adolescent Attitudes toward Serious Mental Illness (ATSMI-AV) (Watson et al. 2005)	24 5-point scaled items	High school students	Grades 9–12	415	US	A
39. Attitudes towards Acute Mental Health Care (ATAMHS) (Baker et al. 2005)	33 7-point scaled items and semantic differentials	Nurses	35–39	140	UK	B
40. Attitudes towards Psychiatry 30 (ATP 30) (Burra et al. 1982)	30 5-point scaled items	Medical students and residents	?	189	Canada	B
41. Latino Scale for Antidepressant Stigma (LSAS) (Interian et al. 2010)	7 3-point scaled items	Latino primary care patients	≥18	200	US	B
42. Stigma Concerns about Mental Health Care (SCMHC) (Interian et al. 2010)	3-point scaled items	Latino primary care patients	≥18	200	US	B
43. Stigma Scale for Receiving Psychological Help (SSRPH) (Komiya et al. 2000)	5 3-point scaled items	College students	M = 18.4 ± 1.32	311	US	B
44. Psychiatric Skepticism scale (PSS) (Swami & Furnham, 2011)	16 7-point scaled items	General public	M = 37.55 ± 14.67	564	UK	B
45. Perceptions of Stigmatisation by Others for Seeking Help (PSOSH) (Vogel et al. 2009)	21 5-point scaled items	College students	?	985; 842; 506; 144; 130	US	B
46. British Omnibus National Survey (ONS) (Kobau et al. 2010)	11 5-point scaled items	Community members	≥18	5251	US	A & B
47. Changing mind (Svensson et al. 2011)	56 5-point scaled items	Nursing students	20–25	51	Sweden	A & B
48. Mental Illness: Clinician's Attitudes (MICA) (Kassam et al. 2010)	16 6-point scaled items	Medical students, psychiatry trainees,	M = 22.4–22.9	23–188	UK	A & B
MICA version 4 (Gabbidon et al. 2013a, b)	16 6-point scaled items	Nursing students	M = 25.56 ± 7.29	191	UK	A & B
49. Libertarian Mental Health Ideology scale (LMHIS) (Nevid & Morrison, 1980)	39 5-point scaled items	Mental health professionals and students	M = 34	227	US	A & B
50. Depression Attitude Questionnaire (DAQ) (Botega et al. 1992)	20 items on a 100 mm visual analogue scale	General practitioners	M = 41 ± 7.4	72	UK	A & B
DAQ (Haddad et al. 2007)	20 items on a 100 mm visual analogue scale	Nurses and home care staff	M = 44.7 ± 9.3	189	UK	A & B



51. DAQ revised (Scheerder <i>et al.</i> 2009)	24 5-point scaled items	Pharmacists	M = 45.2 ± 11.1	200	Belgium	A & B
52. R-DAQ (Haddad <i>et al.</i> 2015)	22 5-point scaled items	Health care providers	?	1193	UK	A & B
53. Opinions about Mental Illness (Cohen & Struening, 1962)	70 6-point scaled items	Health care providers and hospital staff	?	1194	US	A & B
OMI (Madianos <i>et al.</i> 1987)	51 6-point scaled items	Community members	M = 40.9 ± 13.1	1574	Greece	A & B
OMI (Struening & Cohen, 1963)	51 6-point scaled items	Health care providers and hospital staff	?	1200	US	A & B
54. OMI Chinese (Ng & Chan, 2000)	45 6-point scaled items	Secondary students	M = 15.04 ± 1.18	117	China	A & B
55. Community Attitudes towards Mental Illness (CAMI) (Taylor & Dear, 1981)	40 9-point scaled items	Community members	?	1090	Canada	A & B
CAMI (Granello <i>et al.</i> 1999)	40 9-point scaled items	Undergraduate students	18–40	102	US	A & B
CAMI (Granello & Pauley, 2000)	40 9-point scaled items	Undergraduate students	M = 20.54 ± 2.30	53	US	A & B
CAMI (Hinkelman & Granello, 2003)	40 9-point scaled items	Undergraduate students	18–30	86	US	A & B
CAMI (Morris <i>et al.</i> 2011)	40 9-point scaled items	Nurses	M = 40 ± 10	858	6 European countries	A & B
CAMI Chinese (Sevigny <i>et al.</i> 1999)	40 9-point scaled items	Mental health professionals	?	100	China	A & B
CAMI (Wolff <i>et al.</i> 1996)	40 9-point scaled items	General public	≥18	192	UK	A & B
56. CAMI revised (Brockington <i>et al.</i> 1993)	31 5-point scaled items	Community members	≥15	2000	Canada	A & B
57. Mental Health Attitude Survey for Police (MHASP) (Clayfield <i>et al.</i> 2011)	37 3/4-point scaled items	Police officers	M = 41.34 ± 9.09	394	US	A & B
58. CAMI Swedish (Högberg <i>et al.</i> 2008)	20 6-point scaled items	Student nurses	M = 27.9 ± 7.5	256	Sweden	A & B
59. CAMI/FABI (20 item) (Svensson <i>et al.</i> 2011)	20 5-point scaled items	Nursing students	20–25	51	Sweden	A & B
60. Relatives' opinions toward Schizophrenia (ROS) (Magliano <i>et al.</i> 1999)	28 10-point scaled items	Relatives of people with mental illness	M = 55.9 ± 14.8	103	Italy	A & B
61. R-AQ (Pinto <i>et al.</i> 2012)	9 7-point scaled items	High school students	M = 20.15 ± 6.33; M = 20.50 ± 5.87	210	US	A & B
62. Attitudes towards Depression and Its Treatment (ATDT) (Gabriel & Violato, 2010)	27 5-point scaled items	Patients with depression; mental health experts	M = 43; 52 ± 11.6	63; 12	Canada	A & B
63. ATDT revised (Isaac <i>et al.</i> 2012)	25 5-point scaled items	Community members	M = 32.2 ± 12.9	203	Australia	A & B
64. General Attitude Questionnaire (GAQ) (Lam <i>et al.</i> 2005)	Five items measured on 0 to 100 visual-analogue scale	Community members	M = 41.35	110	UK	A & B

Continued

Table 1. Continued

Measurement tools	Response options	Population	Age	Sample	Country	Content
65. Endorsed and Anticipated Stigma Inventory (EASI) (Vogt <i>et al.</i> 2014)	40 5-point scaled items	Military personnel and veterans	M = 37.52 ± 9.99	702	US	A & B
66. emotional reactions (ER) (Angermeyer & Matschinger, 2003)	9 5-point scaled items	Community members	≥18	5025	German	C
67. Affective Reaction scale (Penn <i>et al.</i> 1994)	10 7-point scaled items	College students	?	329	US	C
68. Attribution questionnaire (AQ) (Corrigan <i>et al.</i> 2002)	27 9-point scaled items	College students	M = 26.3 ± 12.2	213	US	A & C
AQ (Brown, 2008)	27 9-point scaled items	College students	M = 19.2 ±	774	US	A & C
AQ (Corrigan <i>et al.</i> 2003)	27 9-point scaled items	College students	M = 25.33 ± 8.77	518	US	A & C
AQ (Corrigan <i>et al.</i> 2004)	27 9-point scaled items	College students	M = 25.7 ± 9.5	54	US	A & C
AQ-27-Italian (Luca Pingani <i>et al.</i> 2012)	27 9-point scaled items	Relatives of college students	M = 40.15 ± 16.36	214	Italy	A & C
69. Consumer Experiences of Stigma Questionnaire (CESQ) -7 items (Bagley & King, 2005)	7 5-point scaled items	Patients with mental illness	M = 42.2	83	UK	D
70. CESQ – 9 items (Świtaj <i>et al.</i> 2013)	7 5-point scaled items	Patients with mental illness	M = 41.5	509	Poland	D
71. Rejection experience (RE) – Swedish (Bjorkman <i>et al.</i> 2007)	11 5-point scaled items	Patients with mental illness	M = 46	40	Sweden	D
72. Personal rejection scale/RE (Moses, 2009)	Six yes/no questions	Youth with mental illness	12–18	60	US	D
73. Discrimination and Stigma scale (DISC) (Brohan <i>et al.</i> 2013)	22 4-point scaled items	Patients with mental illness	M = 41.2 ± 10.9	86	UK	D
74. DISC revised (Thornicroft <i>et al.</i> 2009)	32 7-point scaled items plus four interview questions	People with schizophrenia	M = 39.2 ± 11.32	732	27 countries	D
75. Questionnaire on Anticipated Discrimination (QUAD) (Gabbidon <i>et al.</i> 2013a, b)	17 4-point scaled items	People with mental illness	M = 54 ± 12.69	117	UK	D
76. 15 item Stigma Questionnaire (SQ) (Gibbons <i>et al.</i> 2012)	15 5-point scaled items	People with mental illness	M = 45.7 ± 12; 46.9 ± 16.7	89; 33	Canada	D
77. Harvey stigma scale (Harvey SS) (Harvey, 2001)	18 5-point scaled items	College students	M = 24.07 ± 7.34	197	US	D
78. Harvey SS revised (Bagley & King, 2005)	15 five-point scaled items	Patients with mental illness	M = 42.2	83	UK	D

79. Link's Rejection experiences (Link RE) (Link <i>et al.</i> 1997)	12 multiple choice items	Men with mental illness	M = 34	84	US	D
80. Stigmatising experiences scale (SES) (Stuart <i>et al.</i> 2005)	17 items on prevalence and frequency of stigma experience	People with mental illness	20–79; median = 46	88	Canada	D
81. Self-stigma of depression scale (SSDS) (Barney <i>et al.</i> 2010)	16 5-point scaled items	Community members	M = 50.9	1312	Australia	E
82. Self-stigma of mental illness (SSMI) (Corrigan <i>et al.</i> 2006)	60 9-point scaled items	People with mental illness	M = 41.8 ± 9.6; 44.5 ± 8.5	54; 60	US	E
83. SSMI short form (Corrigan <i>et al.</i> 2006, Corrigan <i>et al.</i> 2012)	20 9-point scaled items	People with mental illness	M = 44.5; 27.8; 35.1; 44.8	71; 60; 30; 85	US, German, Switzerland	E
84. Depression Self-Stigma scale (DSSS) (Kanter <i>et al.</i> 2008)	32 7-point scaled items	Undergraduates and community members	M = 20.93 ± 3.38; 38 ± 13.76	391	US	E
85. The Stigma Inventory for Mental Illness (SIMI) (Karidi <i>et al.</i> 2014)	12 5-point scaled items	Patients with schizophrenia	M = 39.7 ± 9.4	100	Greece	E
86. Link's Secrecy (Link <i>et al.</i> 1989)	8 6-point scaled items	Community members and people with mental illness	M = 32.71–40.29	429; 164	US	E
87. Link's Secrecy (Link <i>et al.</i> 1991)	5 6-point scaled items	People with mental illness	?	152	US	E
88. Link's Withdrawal (Link <i>et al.</i> 1989)	4 6-point scaled items	Community members and people with mental illness	M = 32.71–40.29	429; 164	US	E
89. Link's Withdrawal (Link <i>et al.</i> 1991)	7 6-point scaled items	People with mental illness	?	152	US	E
90. Affiliate self-stigma (ASS) (Mak & Cheung, 2008)	22 4-point scaled items	Caregivers of people with intellectual disability and mental illness	M = 42.81 ± 5.41; 54.21 ± 13.20	210; 108	China	E
91. Self-Stigma scale (SSS) (Mak & Cheung, 2010)	9 4-point scaled items	People with mental illness, recent immigrants	M = 40.07 ± 10.16; 33.98 ± 6.31;	175; 110;	China	E
Self-Stigma Scale-Short (SSSS) (Wu <i>et al.</i> 2015)	9 4-point scaled items	People with mental illness	M = 40.53 ± 10.38; 46.52 ± 11.29	161; 189	China	E
92. Child stigma scale (CSS)/self-stigma (Moses, 2009)	5 4-point Likert scaled items	Youth with mental illness	12–18	60	US	E
93. Secrecy scale (SES) (Moses, 2009)	Seven items	Youth with mental illness	12–18	60	US	E

Continued

Table 1. Continued

Measurement tools	Response options	Population	Age	Sample	Country	Content
94. Internalised stigma of mental illness (ISMI) (Ritsher <i>et al.</i> 2003)	29 4-point scaled items	People with mental illness	M = 49.5 ± 8.7	127	US	E
ISMI (Brohan <i>et al.</i> 2011)	29 4-point scaled items	People with bipolar disorder or depression	M = 45.67 (SD = 12.81)	1182	13 European countries	E
ISMI (Ritsher & Phelan, 2004)	29 4-point scaled items	People with mental illness	M = 51 ± 10	82	US	E
ISMI Chinese (Chang <i>et al.</i> 2014)	29 4-point scaled items	People with mental illness	M = 43.76 ± 11.27	347	China	E
ISMI Arabic (Kira <i>et al.</i> 2015)	29 4-point scaled items	Arab refugees with mental illness	M = 39.66 ± 11.45	330	US	E
ISMI (Lien <i>et al.</i> 2014)	29 4-point scaled items	People with mental illness	M = 43.6 ± 11.76	160	China	E
ISMI (Sibitz <i>et al.</i> 2011a, b)	29 4-point scaled items	People with schizophrenia	M = 37.3 ± 11.9	157	Austria	E
ISMI (Sibitz <i>et al.</i> 2011a, b)	29 4-point scaled items	People with schizophrenia	M = 37.3 ± 11.9	157	Austria	E
ISMI (Sorsdahl <i>et al.</i> 2012)	29 4-point scaled items	Members of depression and anxiety organisation	M = 37 ± 11.3	142	South Africa	E
95. ISMI revised (Assefa <i>et al.</i> 2012)	24 4-point scaled items	People with schizophrenia	M = 33.3 ± 8.9	212	Ethiopia	E
96. ISMI short (Boyd & Otilingam, 2014)	10 4-point scaled items	People with mental illness	M = 49.5; 49.6	127; 760	US	E
97. ISMI (Parent) (Zisman-Ilani <i>et al.</i> 2013)	17 4-point scaled items	Parents of people with mental illness	M = 58.46 ± 4.71	194	Israel	E
98. Self-Stigma of Seeking Psychological Help (SSOSH) (Vogel <i>et al.</i> 2006)	10 5-point scaled items	College students	?	583; 470; 546; 217; 655	US	E
99. Personal stigma scale (PESS) (Schneider <i>et al.</i> 2011)	26 5-point scaled items	People with mental illness	?	243	UK	D & E
100. Stigma of Depression scale (SDS) (Vega <i>et al.</i> 2010)	Seven items	Latino people with depression	M = 50.6 ± 11.3	200	US	A & E
101. Stigma scale (SS) (King <i>et al.</i> 2007)	28 5-point scaled items	Patients with mental illness	M = 42.9 ± 12.4	109	UK	A, C, D

A: Stigma against mental illness or the mentally ill; B: stigma against help-seeking, treatment, mental health institution or psychiatry; C: Emotional responses to mental illness; D: Experienced stigma; E: self-stigma; ?: not reported.

above ( $n=8$ ). Most studies took place in developed countries with the USA as the most studied site ( $n=44$ ), followed by the UK ( $n=21$ ), Canada ( $n=8$ ) and China ( $n=8$ ). The rest of the studies were conducted in 19 different countries.

### Methodological study quality

Table 2 summarises the study quality as: 'excellent', 'good', 'fair' or 'poor'. Each study demonstrated mixed quality from 'poor' to 'good', when addressing different measurement properties of a tool, except one study on the Generalized anxiety stigma scale (GASS) demonstrating 'good' or 'excellent' study quality for all measurement properties assessed (Griffiths *et al.* 2011).

A total of five studies met criteria for 'excellent' quality. These are studies measuring the internal consistency of Stigma-Devaluation scale (Dalky, 2012), the construct and structural validity of GASS (Griffiths *et al.* 2011), as well as the content validity of Opening Minds Scale for Health Care Providers, Self-stigma scale and the revised Discrimination and stigma scale (Thornicroft *et al.* 2009; Mak & Cheung, 2010; Kassam *et al.* 2012).

'Good' quality studies were mostly those measuring internal consistency ( $n=67$ ) (Table 2), followed by five studies on the content validity, one study on test-retest reliability, one study on hypothesis testing (construct validity) and one study on structural validity.

Studies of 'fair' quality were found in most studies evaluating structural validity (89 out of 93), construct validity (hypothesis testing) (85 out of 92), test-retest reliability (38 out of 45), as well as in most studies evaluating cross-cultural validity (three out of four), and all studies ( $n=7$ ) evaluating criterion validity. We further identified studies of 'fair' quality in some studies evaluating internal consistency ( $n=5$ ) and content validity ( $n=8$ ).

No studies on structural validity and criterion validity were identified as of 'poor' quality, however the only two studies [86, 111] (Kassam *et al.* 2010; Modgill *et al.* 2014) on the responsiveness of related tools were rated as 'poor'. We also found some studies with 'poor' quality in evaluating: the internal consistency ( $n=36$ ), content validity ( $n=10$ ), test-retest reliability ( $n=5$ ), construct validity (hypothesis testing) ( $n=5$ ) and cross-cultural validity ( $n=1$ ).

### Level of evidence on the overall quality of measurement properties of stigma tools

As described in previous sections, the study quality (Excellent, Good, Fair or Poor) and the quality of measurement property (+, -, +/- or ?) were combined to determine the level of evidence as: strong (S) (+++ or ---), moderate (M) (++ or --), limited (L) (+ or -), conflicting (C) (+/-), or unknown (U) (x), as

shown in Table 2. The quality of each measurement property helped to determine the direction of the level of evidence of overall quality as positive (+) or negative (-) and their ratings were presented in Table 2 as well.

We found strong evidence (+++) among three tools: the content validity of the revised Discrimination and stigma scale (Thornicroft *et al.* 2009) and Self-stigma scale (Mak & Cheung, 2010); the internal consistency, structural validity (factor analysis) and construct validity of the GASS (Griffiths *et al.* 2011). Moderate level of evidence (M(++); M(--)) were mostly the internal consistency of related tools (55 tools in 63 studies), as well as the content validity of five tools (Table 2). We further found limited level of evidence (L(+); L(-)) for construct validity of 55 tools in 68 studies, structural validity of 46 tools in 56 studies, test-retest reliability of 23 tools in 29 studies, content validity of eight tools, criterion validity of seven tools, and internal consistency of one tool (Table 2).

We identified conflicting (C(+/-)) evidence for the test-retest reliability of nine tools, the internal consistency of six tools, the construct validity of five tools, and the structural validity of three tools (Table 2). We were unable to determine the level of evidence for a number of measurement properties (U(x)) of some tools due to the lack of information provided. This includes the internal consistency of 29 tools (37 studies), structural validity of 25 tools (26 studies), content validity of 11 tools, construct validity of 11 tools, test-retest reliability of four tools and responsiveness of two tools. There are also four tools addressing cross-cultural validity rated as (U(x)) because the COSMIN checklist has not developed criteria for the quality of this property.

Of 101 tools, 12 met the criteria of limited, moderate or strong positive level of evidence on all their assessed measurement properties (highlighted with \*\* in Table 2), and 69 tools reached these levels of evidence for some of their measurement properties. None of the measurement properties for the rest of the 20 tools (highlighted with ?? in Table 2) reached at least the minimum acceptable level of evidence (+).

### Discussion

This review is the first of its kind to investigate the quality of studies containing tools evaluating stigma against mental illness, and the level of evidence of overall quality of measurement properties. As indicated above, a total of 81 tools met the criteria of minimum acceptable, preferred, or ideal level of evidence with positive ratings for all or some of their measurement properties. These results may be useful for researchers and community members to consider for application in practice.

**Table 2.** Methodological quality of included studies and the quality of each measurement property

Measurement tools	Internal consistency	Reliability	Content validity	Structural validity	Hypothesis testing (construct validity)	Cross-cultural validity	Criterion validity	Responsiveness
1. Self-reported prejudiced attitudes (SRPA) (Andersson <i>et al.</i> 2010)	G; +; M(++)			F; ?; U(x)				
2. Personal attributes (PR) (Angermeyer, Matschinger, 2003)				F; +; L(+)	F; ?; U (x)			
3. Labelling (Angermeyer & Matschinger, 2003)	P; +; U(x)				F; +; L(+)			
4. Peer Mental Health Stigmatization scale (PMHSS) (McKeague <i>et al.</i> 2015)	G; +; M(++)	F; -; L(-)	P; -; U(x)	F; +; L(+)	F; +; L(+)			
5. <sup>??</sup> Knowledge Test of Mental Illness (KT) (Michaels & Corrigan, 2013)		F; +/-; C(+/-)		F; -; L(-)			F; -; L(-)	
6. Day's Mental Illness scale (DMIS) (Day <i>et al.</i> 2007)	G; +; M(++)		P; ?; U (x)	F; +; L(+)	F; +; L(+)			
7. <sup>??</sup> EMIC (Chowdhury <i>et al.</i> 2000)	P; +; U (x)	P; +; U (x)						
8. <sup>**</sup> Employer Attitudes towards Mental Illness questionnaire (EAQ) (Diksa & Rogers, 1996)	G; +; M(++)		F; +; L(+)	F; +; L(+)				
9. Attitudes of Nursing Staff towards Co-Workers Returning from Psychiatric and Physical Illnesses (ANCW) (Glozier <i>et al.</i> 2006)	P; +; U (x)		F; +; L(+)		F; +; L(+)			
10. Depression Stigma scale (DSS)/PPSS (Griffiths <i>et al.</i> 2004)	P; +; U (x)	F; +; L(+)						
DSS (Griffiths <i>et al.</i> 2008)	G; +; U (x)			F; -; L(-)	F; +; L(+)			
11. <sup>??</sup> Depression Stigma scale/DSS revised (Gulliver <i>et al.</i> 2012)	P; +; U (x)	P; +; U (x)						
12. Generalized Anxiety Stigma scale (GASS) (Griffiths <i>et al.</i> 2011)	E; +; S(+++)	G; -; M(--)		E; +; S(+++)	E; +; S(+++)			
13. <sup>??</sup> GASS revised (Gulliver <i>et al.</i> 2012)	P; +; U (x)	P; +; U (x)						
14. Test of Knowledge About ADHD (KADD) (Hepperlen <i>et al.</i> 2002)	G; +; M(++)			F; -; L(-)				
15. <sup>**</sup> Beliefs toward Mental Illness (BMI) (Hirai & Clum, 2000)	G; +; M(++)			F; +; L(+)	F; +; L(+)			

16. Opening Minds Scale for Health Care Providers (OMS-HC) (Kassam <i>et al.</i> 2012)	G; +; M(++)	F; -; L(-)	E; +; S(+++)	F; -; L(-)	
OMS-HC (Modgill <i>et al.</i> 2014)	G; +; M(++)			F; -; L(-)	P; ?; U (x)
17. ADHD Stigma Questionnaire (ASQ) (Kellison <i>et al.</i> 2010)	G; +; M(++)	F; +/-; C(+/-)		F; ?; U(x)	F; +; U(x)
ASQ (Bell <i>et al.</i> 2011)	G; +; M(++)			F; ?; U(x)	F; ?; U(x)
18. ADHD stigma (Fuernmaier <i>et al.</i> 2012)	G; +; M(++)			F; +; L(+)	F; ?; U(x)
19. Stigma scale/SS Chinese version (Ho <i>et al.</i> 2015)	G; +; M(++)			F; +; L(+)	F; -; L(-)
20. Self-Esteem and Stigma Questionnaire (SE/SQ) (Hayward <i>et al.</i> 2002)	F; +; L(+)	F; -; L(-)		F; ?; U (x)	F; -; L(-)
21. Devaluation-Discrimination tool (DD) (Link, 1987)	P; +; U(x)				F; +; C(+/-)
Perceived DD (Bjorkman <i>et al.</i> 2007)	P; +; U(x)				F; ?; C(+/-)
Perceived DD (depression) (Interian <i>et al.</i> 2010)	G; -; U(x)			F; +; L(+)	F; -; C(+/-)
22. Depression is a Matter of Will (Aromaa <i>et al.</i> 2010)	G; -; M(--)			F; -; L(-)	F; +; L(+)
23. Public stigma/PDD (Moses, 2009)	G; +; M(++)	F; +; L(+)		F; ?; U(x)	F; +; L(+)
24. Perceived dangerousness (PD) (Link <i>et al.</i> 1987)	P; +; U(x)				F; +; L(+)
25. PD/Link Stigma scale (Bagley & King, 2005)	P; +; U(x)				F; +; L(+)
26. <sup>??</sup> Dangerousness scale (Penn <i>et al.</i> 1994)	P; +; U(x)				F; -; L(-)
27. <sup>??</sup> Social distance (SD) (Link <i>et al.</i> 1989)	P; +; U(x)				F; +; C(+/-)
SD (Penn <i>et al.</i> 1994)	P; +; U(x)				F; -; C(+/-)
28. Social distance revised (Interian <i>et al.</i> 2010)	G; +; M(++)			F; +; L(+)	F; -; L(-)
29. Reported and Intended Behaviour scale (RIBS) (Evans-Lacko <i>et al.</i> 2013)	P; +; U(x)	F; +; L(+)			
RIBS (Evans-Lacko <i>et al.</i> 2012)	P; +; U(x)	F; +; L(+)			
RIBS (Evans-Lacko <i>et al.</i> 2011)	P; +; U(x)	F; +; L(+)	G; +; M(++)		
RIBS (Friedrich <i>et al.</i> 2013)	P; +; U(x)	F; +; L(+)			
RIBS Japanese version (Yamaguchi <i>et al.</i> 2014)	G; +; U(x)	P; +; L(+)		F; ?; U(x)	F; +; L(+)
30. <sup>??</sup> Social Contact scale (SD version) (Jackson & Heatherington, 2006)	G; +/-; C(+/-)			F; ?; U(x)	
31. Social Supports Acceptance scale (SSAS) (Mansouri & Dowell, 1989)	P; +; U(x)				F; +; L(+)

Continued

Table 2. Continued

Measurement tools	Internal consistency	Reliability	Content validity	Structural validity	Hypothesis testing (construct validity)	Cross-cultural validity	Criterion validity	Responsiveness
32. **Attitudes to Mental Illness Questionnaire (AMID) (Luty et al. 2006)		F; +; L(+)		F; +; L(+)				
33. **Attitudes to Severe Mental Illness (ASMI) (Madianos et al. 2012)	G; +; M(++)	F; +; L(+)	F; +; L(+)	F; +; L(+)	F; +; L(+)			
34. Attitudes Toward Social Competence and Integration of People with Mental Illness (ASCI) (Minnebo & Acker, 2004)	G; +; M(++)			P; ?; U(x)				
35. ??Client Attitude Questionnaire (CAQ) (Morrison & Becker, 1975)		P; +; U(x)						
36. Beliefs and attitudes toward people diagnosed with psychosis (BAP) (Serra et al. 2013)	G; +; M(++)			F; ?; U(x)				
37. **Devaluation of consumers family scale (DCFS) (Struening et al. 2001)	G; +; M(++)			F; +; L(+)	F; +; L(+)			
Stigma-Devaluation scale (Dalky, 2012)	E; +; M(++)		F; +; L(+)	E; ?; L(+)		P; N/A; U(x)		
38. **Adolescent Attitudes toward Serious Mental Illness (ATSMI = AV) (Watson et al. 2005)				G; +; M(++)				
39. Attitudes towards Acute Mental Health Care (ATAMHS) (Baker et al. 2005)	F; +; L(+)			F; -; L(-)				
40. Attitudes towards Psychiatry 30 (ATP 30) (Burra et al. 1982)	F; ?; U (x)	F; +/-; C(+/-)	P; ?; U (x)	F; -; L(-)	F; +; L(+)			
41. Latino Scale for Antidepressant Stigma (LSAS) (Interian et al. 2010)	G; -; M(--)			F; +; L(+)	F; -; L(-)			
42. Stigma Concerns about Mental Health Care (SCMHC) (Interian et al. 2010)	G; +/-; C(+/-)			F; +; L(+)	F; -; L(-)			
43. **Stigma Scale for Receiving Psychological Help (SSRPH) (Komiya et al. 2000)	G; +; M(++)			F; +; L(+)	F; +; L(+)			
44. **Psychiatric Skepticism scale (PSS) (Swami & Furnham, 2011)	G; +; M(++)			F; +; L(+)	F; +; L(+)			



45. Perceptions of Stigmatisation by Others for Seeking Help (PSOSH) (Vogel <i>et al.</i> 2009)	G; +; M(++)	F; +; L(+)	P; ?; U(x)	F; +; L(+)	F; +; L(+)	F; +; L(+)
46. <sup>??</sup> British Omnibus National Survey (ONS) (Kobau <i>et al.</i> 2010)	G; -; M(--)			F; -; L(-)	F; -; L(-)	
47. <sup>??</sup> Changing mind (Svensson <i>et al.</i> 2011)	P; -; U(x)	F; -; L(-)				
48. Mental Illness: Clinician's Attitudes (MICA) (Kassam <i>et al.</i> 2010)	G; +; M(++)	F; +; L(+)	G; +; L(+)	F; +; L(+)	F; +/-; C(+/-)	P; ?; U (x)
MICA version 4 (Gabbidon <i>et al.</i> 2013a, b)	G; +; M(++)		F; +; L(+)	F; +; L(+)	F; -; C(+/-)	
49. Libertarian Mental Health Ideology scale (LMHIS) (Nevid & Morrison, 1980)	G; +; M(++)	F; +; L(+)	G; +; M(++)	F; -; L(-)	F; +; L(+)	
50. <sup>??</sup> Depression Attitude Questionnaire (DAQ) (Botega <i>et al.</i> 1992)				F; -; L(-)		
DAQ (Haddad <i>et al.</i> 2007)				F; -; L(-)	F; -; L(-)	
51. <sup>??</sup> DAQ revised (Scheerder <i>et al.</i> 2009)	F; +/-; C(+/-)			F; -; L(-)		
52. R-DAQ (Haddad <i>et al.</i> 2015)	G; +; M(++)	F; -; L(-)	G; +; M(++)	F; -; L(-)		
53. Opinions about Mental Illness (Cohen & Struening, 1962)				F; +; L(+)	F; ?; U (x)	
OMI (Madianos <i>et al.</i> 1987)			F; +; L(+)	F; +; L(+)		
OMI (Struening & Cohen, 1963)	G; +/-; C(+/-)			F ?; L(+)		
54. <sup>??</sup> OMI Chinese (Ng & Chan, 2000)	P; +; U (x)			F; ?; U (x)		
55. Community Attitudes towards Mental Illness (CAMI) (Taylor & Dear, 1981)	G; +/-; C(+/-)		P; ?; U(x)	F; ?; C(+/-)	F; +; L(+)	
CAMI (Granello <i>et al.</i> 1999)	P; +/-; C(+/-)				P; +/-; L(+)	
CAMI (Granello & Pauley, 2000)	P; +/-; C(+/-)				P; +; L(+)	
CAMI (Hinkelman & Granello, 2003)	G; +/-; C(+/-)				F; +; L(+)	
CAMI (Morris <i>et al.</i> 2011)				F; ?; C(+/-)		
CAMI Chinese (Sevigny <i>et al.</i> 1999)				F; -; C(+/-)	F; +; L(+)	
CAMI (Wolff <i>et al.</i> 1996)				F; -; C(+/-)	F; +; L(+)	
56. <sup>**</sup> CAMI revised (Brockington <i>et al.</i> 1993)				F; +; L(+)	F; +; L(+)	
57. Mental Health Attitude Survey for Police (MHASP) (Clayfield <i>et al.</i> 2011)	G; +; M(++)			F; -; L(-)	P; +; U(x)	
58. CAMI Swedish (Högberg <i>et al.</i> 2008)	G; +; M(++)			F; ?; U(x)		F; N/A; U(x)
59. <sup>??</sup> CAMI/FABI (20 item) (Svensson <i>et al.</i> 2011)	P +; U(x)	F; -; L(-)				

Continued

Table 2. Continued

Measurement tools	Internal consistency	Reliability	Content validity	Structural validity	Hypothesis testing (construct validity)	Cross-cultural validity	Criterion validity	Responsiveness
60. Relatives' opinions toward Schizophrenia (ROS) (Magliano <i>et al.</i> 1999)	G; -; M(--)	F; +/-; C(+/-)		F; +; L(+)	P; -; U(x)			
61. R-AQ (Pinto <i>et al.</i> 2012)	G; +; M(++)			F; -; L(-)				
62. Attitudes towards Depression and Its Treatment (ATDT) (Gabriel & Violato, 2010)	G; +; M(++)		P; ?; U (x)	F; +; L(+)				
63. <sup>??</sup> ATDT revised (Isaac <i>et al.</i> 2012)	G; -; M(--)			F; ?; U (x)				
64. <sup>??</sup> General Attitude Questionnaire (GAQ) (Lam <i>et al.</i> 2005)	P; +; U(x)	F; +/-; C(+/-)						
65. Endorsed and Anticipated Stigma Inventory (EASI) (Vogt <i>et al.</i> 2014)	F; ?; U(x)		P; ?; U(x)	F; ?; U(x)	F; +; L(+)			
66. emotional reactions (ER) (Angermeyer & Matschinger, 2003)				F; +; L(+)	F; ?; U (x)			
67. <sup>??</sup> Affective Reaction scale (Penn <i>et al.</i> 1994)	P; +; U(x)				F; -; L(-)			
68. Attribution questionnaire (AQ)-27 (Corrigan <i>et al.</i> 2002)				F; ?; C(+/-)	F; ?; C(+/-)			
AQ (Brown, 2008)	G; +/-; C(+/-)	F; +; L(+)		F; +; C(+/-)	F; +/-; C(+/-)			
AQ (Corrigan <i>et al.</i> 2003)	P; +; C(+/-)				F; +; C(+/-)			
AQ (Corrigan <i>et al.</i> 2004)		F; +/-; L(+)						
AQ-27-Italian (Luca Pingani <i>et al.</i> 2012)	G; +; C(+/-)	F; +; L(+)		F; ?; C(+/-)	F; +; C(+/-)	F; N/A; U (x)		
69. Consumer Experiences of Stigma Questionnaire (CESQ) -7 items (Bagley & King, 2005)	P; +; U(x)				F; +; L(+)			
70. CESQ - 9 items (Świtaj <i>et al.</i> 2013)	G; +; M(++)			F; ?; U (x)	F; +; L(+)	F; N/A; U (x)		
71. <sup>??</sup> Rejection experience (RE) - Swedish (Bjorkman <i>et al.</i> 2007)	P; +; U(x)				F; ?; U(x)			
72. Personal rejection scale/RE (Moses, 2009)	G; +; M(++)	F; +; L(+)		F; ?; U (x)	F; +; L(+)			
73. <sup>??</sup> Discrimination and Stigma scale (DISC) (Brohan <i>et al.</i> 2013)	P; +; U(x)	F; +/-; C(+/-)			F; -; L(-)			
74. <sup>**</sup> DISC revised (Thornicroft <i>et al.</i> 2009)			E; +; S(+++)					

75. Questionnaire on Anticipated Discrimination (QUAD) (Gabbidon <i>et al.</i> 2013a, b)	G; +; M(++)	F; +; L(+)		F; -; L(-)	F; -; L(-)	
76. 15 item Stigma Questionnaire (SQ) (Gibbons <i>et al.</i> 2012)	P; +; U (x)	F; +; L(+)	G; ?; U (x)		F; +; L(+)	F; +; L(+)
77. Harvey stigma scale (Harvey SS) (Harvey, 2001)	G; +; M(++)		F; +; L(+)	F; ?; U (x)	F; +; L(+)	
78. Harvey SS revised (Bagley & King, 2005)	P; +; U (x)				F; +; L(+)	
79. Link's Rejection experiences (Link RE) (Link <i>et al.</i> 1997)	G; +; M(++)			F; ?; U(x)		
80. Stigmatizing experiences scale (SES) (Stuart <i>et al.</i> 2005)	P; +; U(x)		G; +; M(++)		F; +; L(+)	
81. Self-stigma of depression scale (SSDS) (Barney <i>et al.</i> 2010)	G; +; M(++)	F; -; L(-)		F; +; L(+)	F; +; L(+)	
82. Self-stigma of mental illness (SSMI) (Corrigan <i>et al.</i> 2006)	P; +/-; U (x)	F; +/-; C(+/-)			F; +; L(+)	
83. SSMI short form (Corrigan <i>et al.</i> 2006, Corrigan <i>et al.</i> 2012)	P; +; U (x)				F; +; L(+)	
84. Depression self-stigma scale (DSSS) (Kanter <i>et al.</i> 2008)	G; +; M(++)			F; ?; U (x)	F; +; L(+)	
85. The Stigma Inventory for Mental Illness (SIMI) (Karidi <i>et al.</i> 2014)	G; +; M(++)	F; +; L(+)	P; ?; U (x)	F; ?; U (x)	P; -; U (x)	F; -; L(-)
86. Link's Secrecy (Link <i>et al.</i> 1989)	G; +; M(++)			F; ?; U(x)	F; +; L(+)	
87. Link's Secrecy (Link <i>et al.</i> 1991)	G; +; M(++)			F; ?; U(x)	F; +; L(+)	
88. Link's Withdrawal (Link <i>et al.</i> 1989)	G; -; M(--)			F; ?; U(x)	F; +; L(+)	
89. Link's Withdrawal (Link <i>et al.</i> 1991)	G; -; M(--)			F; ?; U(x)	F; +; L(+)	
90. Affiliate self-stigma (ASS) (Mak & Cheung, 2008)	G; +; M(++)			F; -; L(-)	F; +; L(+)	
91. Self-stigma scale (SSS) (Mak & Cheung, 2010)	G; +; M(++)		E; +; S(+++)	F; +; C(+/-)	F; +; C(+/-)	
Self-stigma scale-short (SSSS) (Wu <i>et al.</i> 2015)	G; +; M(++)			F; ?; C(+/-)	F; -; C(+/-)	
92. Child stigma scale (CSS)/self-stigma (Moses, 2009)	G; +; M(++)	F; +; L(+)		F; ?; U(x)	F; +; L(+)	
93. Secrecy scale (SES) (Moses, 2009)	G; +; M(++)	F; +; L(+)		F; ?; U(x)	F; +; L(+)	
94. Internalised stigma of mental illness (ISMI) (Ritsher <i>et al.</i> 2003)	G; +; M(++)	P; +; C(+/-)		F; ?; L(+)	F; +; L(+)	
ISMI (Brohan <i>et al.</i> 2011)	P; +; M(++)				F; +; L(+)	
ISMI (Ritsher & Phelan, 2004)					F; +; L(+)	
ISMI Chinese (Chang <i>et al.</i> 2014)	G; +; M(++)	F; +; C(+/-)		F; ?; L(+)	F; +/-; L(+)	
ISMI Arabic (Kira <i>et al.</i> 2015)	G; +; M(++)			F; +; L(+)	F; -; L(+)	F; -; L(-)

Continued

Table 2. Continued

Measurement tools	Internal consistency	Reliability	Content validity	Structural validity	Hypothesis testing (construct validity)	Cross-cultural validity	Criterion validity	Responsiveness
ISMI (Lien <i>et al.</i> 2014)	G; +; M(++)	F; +/-; C(+/-)		F; +; L(+)	F; +; L(+)			
ISMI (Sibitz <i>et al.</i> 2011a, b)					G; +; L(+)			
ISMI (Sibitz <i>et al.</i> 2011a, b)				F; +; L(+)	F; +; L(+)			
ISMI (Sorsdahl <i>et al.</i> 2012)					F; +; L(+)			
95. **ISMI revised (Assefa <i>et al.</i> 2012)	G; +; M(++)			F; +; L(+)	F; +; L(+)			
96. ISMI short (Boyd & Otilingam, 2014)	P; +; U(x)		F; +; L(+)		F; +; L(+)			
97. ISMI (Parent) (Zisman-Ilani <i>et al.</i> 2013)	G; +; M(++)		P; ?; U(x)	F; +; L(+)				
98. Self-Stigma of Seeking Psychological Help (SSOSH) (Vogel <i>et al.</i> 2006)	G; +; M(++)	F; +; L(+)	P; ?; U(x)	F; +; L(+)	F; +; L(+)		F; -; L(-)	
99. **Personal stigma scale (PESS) (Schneider <i>et al.</i> 2011)	G; +; M(++)			F; +; L(+)				
100. ??Stigma of depression scale (SDS) (Vega <i>et al.</i> 2010)	G; -; M(---)			F; ?; U(x)				
101. Stigma scale (SS) (King <i>et al.</i> 2007)	G; +; M(++)	F; +/-; C(+/-)		F; +; L(+)	F; +; L(+)			

Study quality: E = Excellent, G = Good, F = Fair, P = Poor; Quality of each measurement property: positive rating (+), negative rating (-), indeterminate rating (?), conflicting rating (+/-); Overall level of evidence: Strong (S) (+++ or ---), Moderate (M) (++ or --), Limited (L) (+ or -), Conflicting (C) (+/-), or unknown (U) (x); N/A = Not applicable.

\*\*, 12 tools of which all their measurement properties met the criteria of Limited (+ or -) (minimum acceptable) evidence or above; ??, 20 tools of which no measurement properties met the criteria of minimum acceptable evidence (limited level of evidence) or above.

However, it is a challenge to conclude one tool is better than the other for a number of reasons: (1) included tools contained different items addressing various domains of stigma, even for tools developed under the same theoretical framework; (2) studies evaluated different measurement properties; and (3) study quality and level of evidence varied even in the same study depending on the properties measured. For example, Attitudes to Severe Mental Illness measured general attitudes of the general public and is one of the 12 tools of which all measurement properties reached 'limited' or 'moderate' level of evidence (Madianos *et al.* 2012). Another tool, Reported and Intended Behaviour scale (Evans-Lacko *et al.* 2011) also measured general attitudes of the general public in multiple studies and had mixed level of evidence from 'unknown' (x) to 'moderate' (++). In this circumstance when choosing which tool for application, evidence of each individual property matters and we should also consider whether the purpose of the chosen tool (e.g., the content of the tool, target population, and the setting) is consistent with our actual application, either in developing an anti-stigma intervention or to measure public stigma of mental illness.

Based on the current evidence, we recommend to use the 12 tools with all their evaluated measurement properties reaching at least 'limited' level of evidence or above (highlighted with \*\* in Table), as well as tools reaching these quality levels (limited or above) for at least half of their evaluated measurement properties (Table 2). Yet, we do not recommend tools with negative ratings (---, -- or -) because the statistics of these measurement properties were below the criteria threshold, nor are we confident about the application of tools with conflicting (+/-) or unknown (x) evidence. We also however raise the caveat that future recommendations on the use of these tools may change as we know that the validation of a tool is an ongoing process (Streiner & Norman, 2008) and as more studies are conducted with more appropriate designs, tools that currently do not meet our criteria may do so following further future research.

The finding that there are currently over 100 different stigma measurement tools raises concerns about the overall value of this body of research, as it is simply not possible to come to general considerations about issues related to stigma in mental illness given the use of so many different tools to measure the concept. As such, we were unable to decide which tool is the 'gold standard' in this area and this is probably why only 2 (Vogel *et al.* 2009; Gibbons *et al.* 2012) out of seven studies measuring criterion validity showed significant correlations with the pre-defined 'gold standard' tools. Future research should focus on using a much smaller number of tools, those with the best

psychometric properties to help decrease the uncertainty arising from the application of so many different tools of varying quality. One important step to achieve this goal may be to reconstruct and synthesise various stigma theories and reach consensus on what a measure of stigma against mental illness should entail.

The study characteristics of these included validated tools are consistent with findings from the scoping review (Wei *et al.* 2015) that there are few tools (six tools) assessing people's emotional responses to mental illness. Further, most research was conducted in the USA and it is not known if tools applied this population can be compared with those applied in other countries. Similarly, there are few tools validated among secondary school students ( $n=8$ ) and teachers ( $n=2$ ), indicating a substantial contrast against the fact that most mental disorders onset between the age of 12 and 25 (Kieling *et al.* 2011) and most young people attend school during this period of time.

Measuring stigma against mental illness is challenging because of social desirability bias where people tend to answer questions in a manner that will be viewed favourably by others (Maccoby & Maccoby, 1954). This bias may seriously jeopardise the validity of findings when the tool is applied. We found that only 1 out of the 101 tools addressed this potential bias by applying error-choice response (Hepperlen *et al.* 2002). Future application of stigma tools may need to consider evidence-based approaches to reduce social desirability bias. Some recommended techniques include the integration of social desirability scale assessment into the stigma assessment tool, the application of random response techniques, the addition of disguising of scale intent or an indirect questioning approach (Streiner & Norman, 2008).

Based on our findings and informed by the COSMIN checklist, we also have recommendations for researchers to consider. First, psychometric studies need to obtain an adequate sample size, and address missing items for relevant measurement properties. In addition, checking unidimensionality of items is as important as reporting Cronbach's alpha or KR-20 in deciding the study quality of internal consistency. Further, in examining test-retest reliability, the analysis on the independence of the test administration, the appropriate timing between tests, and the stability of test conditions were often ignored but matter in improving study quality. When assessing content validity, piloting the items in the targeting population ( $\geq 10$ ) for comprehensiveness is equally important as item selection process. In analysing the structural validity/factor analysis, it is essential that researchers report the variances explained by factor analysis to improve study quality. When measuring construct validity, it is suggested that studies formulate hypotheses in advance and pre-

define the direction and the magnitude of the mean difference or correlations of related statistical analysis to ensure the appropriateness of analysis.

It is noted that the most assessed measurement properties were internal consistency, structural and construct validity, while responsiveness was the least studied property and measurement errors were not assessed by included studies. Rising from this analysis is the question of what and how many psychometric properties should be included for psychometric analysis. Although the COSMIN checklist established criteria for nine properties, it is a modular framework that does not require the evaluator to complete analysis of all nine properties. However, informed by the findings from this review, it is reasonable to propose that the validation of a tool should at least analyse whether: the tool items are appropriately related (internal consistency); it is reliable over time (test-retest reliability); and the tool constructs are adequately established (structural and construct validity).

Additionally, when it is applied in culturally different settings, cross-cultural validity has to be evaluated prior to its application. The lack of cross-culturally validated tools (only four tools) makes cross cultural conclusions about stigma against mental illness difficult if not impossible. To address cross-cultural validity, researcher should make sure the culturally adapted tool is an adequate reflection of the original one. This could be achieved through a number of processes, including: multiple forward and backward translations of the tool with a committee to review the final translation; a pre-test of the tool with the target population performed to check cultural relevance; and the hypothesised factor structure tested with confirmatory factor analysis.

### Limitations

Our review is limited in excluding non-English publications (25 non-English potentially relevant citations were identified at the title and abstract screening stages) and therefore may have missed some eligible studies otherwise. Secondly, the COSMIN checklist may not be the most appropriate critical appraisal approach although it is the only available one, because it is originally designed for health status questionnaire.

### Conclusions

This is the first systematic review to investigate the study quality and overall level of evidence of tools evaluating stigma of mental illness. We categorised included tools, and provided rich evidence on the psychometric properties of current stigma measurement tools so that researchers and decision makers can

choose best available tools for use in practice. However, no matter what tools researchers or decision makers choose, it is recommended that researchers continue to validate tools in different settings to ensure that these tools are able to be appropriately used in numerous different contexts and populations.

### Acknowledgements

We would like to acknowledge that this study is supported by Yifeng Wei's Doctoral Research Award – Priority Announcement: Knowledge Translation/Bourse de recherche, issued by the Canadian Institutes of Health Research. Dr McGrath is supported by a Canada Research Chair. In addition, we would like to thank Ms Catherine Morgan and Michelle Xie for their help with data collection and analysis, and the health librarian, Ms Robin Parker, who helped with designing the search strategies of this review.

### Conflict of Interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

### Ethical Standards

An approval by ethics committee was not applicable to this review.

### Availability of Data and Materials

Owing to the large amount of data (risk of bias analysis, quality of each measurement properties for 117 studies), we choose to share it upon audience's requests.

### References

- Andersson HW, Bjørngaard JH, Silje Lill Kaspersen SL, Wang CEA, Skre I, Dahl T (2010). The effects of individual factors and school environment on mental health and prejudiced attitudes among Norwegian adolescents. *Social Psychiatry and Psychiatric Epidemiology* **45**, 569–577. doi: 10.1007/s00127-009-0099-0.
- Angermeyer MC, Matschinger H (2003). The stigma of mental illness: effects of labeling on public attitudes towards people with mental disorder. *Acta Psychiatrica Scandinavica* **108**, 304–309.
- Aromaa E, Tolvanen A, Tuulari J, Wahlbeck K (2010). Attitudes towards people with mental disorders: the psychometric characteristics of a Finnish questionnaire. *Social Psychiatry & Psychiatric Epidemiology* **45**, 265–273. doi: 10.1007/s00127-009-0064-y.

- Assefa D, Shibre T, Asher L, Fekadu A (2012). Internalized stigma among patients with schizophrenia in Ethiopia: a cross-sectional facility-based study. *BMC Psychiatry* **12**, 239. <http://www.biomedcentral.com/1471-244X/12/239>.
- Bagley C, King M (2005). Exploration of three stigma scales in 83 users of mental health services: implications for campaigns to reduce stigma. *Journal of Mental Health* **14**, 343–355. doi: 10.1080/09638230500195270.
- Baker JA, Richards DA, Campbell M (2005). Nursing attitudes towards acute mental health care: development of a measurement tool. *Journal of Advanced Nursing* **49**, 522–529.
- Barney LJ, Griffiths KM, Christensen H, Jorm AF (2010). The self-stigma of depression scale (SSDS): development and psychometric evaluation of a new instrument. *International Journal of Methods in Psychiatric Research* **19**, 243–254. doi: 10.1002/mpr.325.
- Batterham PJ, Griffiths KM, Barney LJ, Parsons A (2013). Predictors of generalized anxiety disorder stigma. *Psychiatry Research* **206**, 282–286.
- Bell L, Long S, Garvan C, Bussing R (2011). The impact of teacher credentials on ADHD stigma perceptions. *Psychology in the Schools* **48**, 184–197. doi: 10.1002/pits.20536.
- Bjorkman T, Svensson B, Lundberg B (2007). Experiences of stigma among people with severe mental illness. Reliability, acceptability and construct validity of the Swedish versions of two stigma scales measuring devaluation/discrimination and rejection experiences. *Nordic Journal Psychiatry* **61**, 332–338. doi: 10.1080/08039480701642961.
- Botega N, Mann A, Blizard R, Wilinson G (1992). General practitioners and depression – first use of the depression attitude questionnaire. *International Journal of Methods in Psychiatric Research* **2**, 169–180.
- Boyd JE, Otilingam PG (2014). Brief version of the internalized stigma of mental illness (ISMI) scale: psychometric properties and relationship to depression, self esteem, recovery orientation, empowerment, and perceived devaluation and discrimination. *Psychiatric Rehabilitation Journal* **37**, 17–23. doi: 10.1037/prj000003517.
- Boyd JE, Adler EP, Otilingam PG, Peters T (2014). Internalized stigma of mental illness (ISMI) scale: a multinational review. *Comprehensive Psychiatry* **55**, 221–231.
- Brockington IF, Hall P, Levings J, Murphy C (1993). The community's tolerance of the mentally ill. *British Journal of Psychiatry* **162**, 93–99.
- Brohan E, Slade M, Clement S, Thornicroft G (2010). Experiences of mental illness stigma, prejudice and discrimination: a review of measures. *BMC Health Services Research* **10**, 80. <http://www.biomedcentral.com/1472-6963/10/80>.
- Brohan E, Gauci D, Sartorius N, Thornicroft G, For the GAMIAN-Europe Study Group (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, Clement S, Rose D, Sartorius N, Slade M, Thornicroft G (2013). Development and psychometric evaluation of the discrimination and stigma scale (DISC). *Psychiatry Research* **208**, 33–40. doi: 10.1016/j.psychres.2013.03.007.
- Brown SA (2008). Factors and measurement of mental illness stigma: a psychometric examination of the attribution questionnaire. *Psychiatric Rehabilitation J* **32**, 89–94. doi: 10.2975/32.2.2008.89.94.
- Burra P, Kalin R, Lechner P, Waldron JJ, Handforth JR, Jarrett FJ, Amara IB (1982). The ATP 30 – a scale for measuring medical students' attitudes of psychiatry. *Medical Education* **16**, 31–38.
- Chang C, Wu T, Chen C, Wang J, Lin C (2014). Psychometric evaluation of the internalized stigma of mental illness scale for patients with mental illnesses: measurement invariance across time. *PLoS ONE* **9**, e98767. doi: 10.1371/journal.pone.0098767.
- Chowdhury AN, Sanyal D, Dutta SK, Banerjee S, De R, Bhattacharya K, Palit S, Bhattacharya P, Monda RK, Weiss MG (2000). Stigma and mental illness: pilot study of laypersons and health care providers with the EMIC in rural west Bengal, India. *International Medical Journal* **7**, 257–260.
- Clayfield JC, Fletcher KE, Grudzinskas AJ Jr. (2011). Development and validation of the mental health attitude survey for police. *Community Mental Health Journal* **47**, 742–751. doi: 10.1007/s10597-011-9384-y.
- Cohen J, Struening EL (1962). Opinions about mental illness in the personnel of two large mental hospitals. *Journal of Abnormal and Social Psychology* **64**, 349–360.
- Corrigan PW, Rowan D, Qreen A, Lundin R, River P, Uphoff-Wasowski K, White K, Kubiak MA (2002). Challenging two mental illness stigmas: personal responsibility and dangerousness. *Schizophrenia Bulletin* **28**, 293–309.
- Corrigan P, Markowitz FE, Watson A, Rowan D, Kubiak MA (2003). An attribution model of public discrimination towards persons with mental illness. *Journal of Health and Social Behavior* **44**, 162–179.
- Corrigan PW, Watson AC, Warpinski AC, Gracia G (2004). Stigmatizing attitudes about mental illness and allocation of resources to mental health services. *Community Mental Health Journal* **40**, 297–307.
- Corrigan PW, Watson AC, Barr L (2006). The self-stigma of mental illness: implications for self-esteem and self-efficacy. *J Social and Clinical Psychology* **25**, 875–884.
- Corrigan PW, Michaels PJ, Vega E, Gause M, Watson AC, Rusch N (2012). Self-stigma of mental illness scale-short form: reliability and validity. *Psychiatry Research* **199**, 65–69. doi: 10.1016/j.psychres.2012.04.009.
- Dalky HF (2012). Arabic translation and cultural adaptation of the stigma-devaluation scale in Jordan. *Journal of Mental Health* **21**, 72–82. doi: 10.3109/09638237.2011.629238.
- Day EN, Edgren K, Eshleman A (2007). Measuring stigma toward mental illness: development and application of the mental illness stigma scale. *Journal of Applied Social Psychology* **10**, 2191–2219.
- Diksa E, Rogers ES (1996). Employer concerns about hiring persons with psychiatric disability: results of the employer attitude questionnaire. *Rehabilitation Counseling Bulletin* **40**, 31–44.
- Evans-Lacko S, Rose D, Little K, Flach C, Rhydderch D, Henderson C, Thornicroft G (2011). Development and psychometric properties of the reported and intended behavior scale (RIBS): a stigma-related behavior measure. *Epidemiology and Psychiatric Sciences* **20**, 263–271.
- Evans-Lacko S, London J, Japhet S, Rusch N, Flach C, Corker E, Henderson C, Thornicroft G (2012). Mass social contact interventions and their effect on mental health

- related stigma and intended discrimination. *BMC Public Health* **12**, 489.
- Evans-Lacko S, Henderson C, Thornicroft G** (2013). Public knowledge, attitudes and behavior regarding people with mental illness in England 2009–2012. *British Journal of Psychiatry* **202**(Suppl.), 51–57. doi: 10.1192/bjp.bp.112.112979.
- Friedrich B, Evans-Lacko S, London J, Rhydderch D, Henderson C, Thornicroft G** (2013). Anti-stigma training for medical students: the education not discrimination project. *British Journal of Psychiatry* **202**(Suppl.), 89–94. doi: 10.1192/bjp.bp.112.114017.
- Fuermaier ABM, Tucha L, Koerts J, Mueller AK, Lange KW, Tucha O** (2012). Measurement of stigmatization towards Adults with Attention Deficit Hyperactivity Disorder. *PLoS ONE* **7**, e51755. doi: 10.1371/journal.pone.0051755.
- Furlan AD, Malmivaara A, Chou R, Maher CG, Deyo RA, Schoene M, Bronfort G, van Tulder MW, Editorial Board of the Cochrane Back, Neck Group** (2015). 2015 updated method guidelines for systematic reviews in the Cochrane Back Review Group. *Spine (Phila Pa 1976)* **40**, 1660–1673. doi: 10.1097/BRS.0000000000001061.
- Gabbidon J, Brohan E, Clement S, Henderson RC, Thornicroft G, MIRIAD Study Group** (2013a). The development and validation of the Questionnaire on Anticipated Discrimination (QUAD). *BMC Psychiatry* **13**, 297. <http://www.biomedcentral.com/1471-244X/13/297>. Accessed 10 June 2016.
- Gabbidon J, Clement S, Nieuwenhuizen A, Kassam A, Brohan E, Norman I, Thornicroft G** (2013b). Mental Illness: Clinicians' Attitudes (MICA) Scale – Psychometric properties of a version for health care students and professionals. *Psychiatry Research* **206**, 81–87. doi: 10.1016/j.psychres.2012.09.028.
- Gabriel A, Violato C** (2010). The development and psychometric assessment of an instrument to measure attitudes towards depression and its treatments in patients suffering from non-psychotic depression. *Journal of Affective Disorders* **124**, 241–249. doi: 10.1016/j.jad.2009.11.009.
- Gibbons C, Dubois S, Morris K, Parker B, Maxwell H, Bédard M** (2012). The development of a questionnaire to explore stigma from the perspective of individuals with serious mental illness. *Canadian Journal of Community Mental Health* **31**, 17–32.
- Glozier N, Hough C, Henderson M, Holland-Elliott K** (2006). Attitudes of nursing staff towards co-workers returning from psychiatric and physical illnesses. *International Journal of Social Psychiatry* **52**, 525–534. doi: 10.1177/00207640060666843.
- Goffman E** (1963). *Stigma: Notes on the Management of Spoiled Identity*. Prentice Hall: Englewood Cliffs, NJ.
- Granello DH, Pauley PS** (2000). Television viewing habits and their relationship to tolerance towards people with mental illness. *Journal of Mental Health Counseling* **22**, 162–175.
- Granello DH, Pauley PS, Carmichael A** (1999). Relationship of the media to attitudes toward people with mental illness. *Journal of Humanistic Counseling* **38**, 98–110.
- Griffiths KM, Christensen H, Jorm AF, Evans K, Groves C** (2004). Effect of web-based depression literacy and cognitive-behavioural therapy interventions on stigmatizing attitudes to depression: randomized controlled trial. *British Journal of Psychiatry* **185**, 342–349. doi: 10.1192/bjp.185.4.342.
- Griffiths KM, Christensen H, Jorm AF** (2008). Predictors of depression stigma. *BMC Psychiatry* **8**, 25. doi: 10.1186/1471-244X-8-25.
- Griffiths KM, Batterham PJ, Barney L, Parsons A** (2011). The generalized anxiety stigma scale (GASS): psychometric properties in a community sample. *BMC Psychiatry* **11**, 184. <http://www.biomedcentral.com/1471-244X/11/184>. Accessed 10 June 2016.
- Gulliver A, Griffiths KM, Christensen H** (2010). Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. *BMC Psychiatry* **10**, 113.
- Gulliver A, Griffiths KM, Christensen H, Mackinnon A, Calear AL, Parsons A, Bennett K, Batterham PJ, Stanimirovic R** (2012). Internet-based interventions to promote mental health help-seeking in elite athletes: an exploratory randomized controlled trial. *Journal of Medical Internet Research* **14**, e69. doi: 10.2196/jmir.1864.
- Haddad M, Walters P, Tylee A** (2007). District nursing staff and depression: a psychometric evaluation of depression attitude questionnaire findings. *International Journal of Nursing Studies* **44**, 447–456.
- Haddad M, Menchetti M, McKeown E, Tylee A, Mann A** (2015). The development and psychometric properties of a measure of clinicians' attitudes to depression: the revised Depression Attitude Questionnaire (R-DAQ). *BMC Psychiatry* **15**, 7. doi: 10.1186/s12888-014-0381-x.
- Harvey RD** (2001). Individual differences in the phenomenological impact of social stigma. *Journal of Social Psychology* **141**, 174–189.
- Hayward P, Wong G, Bright JA, Lam D** (2002). Stigma and self-esteem in manic depression: an exploratory study. *Journal of Affective Disorders* **69**, 61–67.
- Hepperlen TM, Clay DL, Henly GA, Barké CR, Hepperlen MH, Clay DL** (2002). Measuring teacher attitudes and expectations toward students with ADHD: development of the test of knowledge about ADHD (KADD). *Journal of Attention Disorders* **5**, 133–142. doi: 10.1177/108795470200500301.
- Hinkelmean L, Granello DH** (2003). Biological sex, adherence to traditional gender roles, and attitudes toward persons with mental illness: an exploratory investigation. *Journal of Mental Health Counseling* **25**, 259–270.
- Hirai M, Clum GA** (2000). Development, reliability, and validity of the beliefs toward mental illness scale. *Journal of Psychopathology Behavioral Assessment* **22**, 221–236.
- Ho AHY, Potash JS, Fong TCT, Ho VFL, Chen EYH, Lau RHW, Au Yeung FS, Ho RT** (2015). Psychometric properties of a Chinese version of the stigma scale: examining the complex experience of stigma and its relationship with self-esteem and depression among people living with mental illness in Hong Kong. *Comprehensive Psychiatry* **56**, 198–205. doi: 10.1016/j.comppsy.2014.09.016.
- Högberg T, Magnusson A, Ewertzon M, Lützn K** (2008). Attitudes towards mental illness in Sweden: adaptation and development of the Community Attitudes towards Mental Illness questionnaire. *International Journal of Mental*



- Health Nursing* 17, 302–310. doi: 10.1111/j.1447-0349.2008.00552.x.
- Interian A, Ang A, Gara MA, Link B, Rodriguez MA, Vega WA** (2010). Stigma and depression treatment utilization among Latinos: utility of four stigma measures. *Psychiatric Services* 61, 373–379.
- Isaac F, Greenwood KM, Benedetto M** (2012). Evaluating the psychometric properties of the attitudes towards depression and its treatments scale in an Australian sample. *Patient Prefer Adherence* 6, 349–354. doi: 10.2147/PPA.S26783.
- Jackson D, Heatherington L** (2006). Young Jamaicans' attitudes toward mental illness: experimental and demographic factors associated with social distance and stigmatizing opinions. *Journal of Community Psychology* 34, 563–576. doi: 10.1002/jcop.20115.
- Jones EE, Farina A, Hastorf AH, Marcus H, Miller DT, Scott RA** (1984). *Social Stigma: the Psychology of Marked Relationships*. Freeman and Company: New York.
- Kanter JW, Rusch LC, Brondino MJ** (2008). Depression Self-Stigma: a new measure and preliminary findings. *Journal of Nervous and Mental Disease* 196, 663–670. doi: 10.1097/NMD.0b013e318183f8af.
- Karidi MV, Vasilopoulou D, Savvidou E, Vitoratou S, Rabavilas AD, Stefanis CN** (2014). Aspects of perceived stigma: the stigma inventory for mental illness, its development, latent structure and psychometric properties. *Comprehensive Psychiatry* 55, 1620–1625. doi: 10.1016/j.comppsy.2014.04.002.
- Kassam A, Glozier N, Leese M, Henderson C, Thornicroft G** (2010). Development and responsiveness of a scale to measure clinicians' attitudes to people with mental illness (medical student version). *Acta Psychiatrica Scandinavica* 122, 153–161. doi: 10.1111/j.1600-0447.2010.01562.x.
- Kassam A, Papish A, Modgill G, Patten S** (2012). The development and psychometric properties of a new scale to measure mental illness related stigma by health care providers: the opening minds scale for Health Care Providers (OMS-HC). *BMC Psychiatry* 12, 62. <http://www.biomedcentral.com/1471-244X/12/62>. Accessed 14 August 2015.
- Kellison I, Bussing R, Bell L, Garvan C** (2010). Assessment of stigma associated with attention-deficit hyperactivity disorder: psychometric evaluation of the ADHD stigma questionnaire. *Psychiatry Research* 178, 363–369. doi: 10.1016/j.psychres.2009.04.022.
- Kieling C, Baker-Henningham H, Belfer M, Conti G, Ertem I, Omigbodun O, Rohde LA, Srinath S, Ulkuer N, Rahman A** (2011). Child and adolescent mental health worldwide: evidence for action. *Lancet* 378, 1515–1525. doi: 10.1016/S0140-6736(11)60827-1. Epub 2011 Oct 16.
- King M, Dinos S, Shaw J, Watson R, Stevens S, Passetti F, Weich S, Serfaty M** (2007). The stigma scale: development of a standardized measure of the stigma of mental illness. *British Journal of Psychiatry* 190, 248–254. doi: 10.1192/bjp.bp.106.024638.
- Kira IA, Ramaswamy V, Lewandowski L, Mohanesh J, Abdul-Khalek H** (2015). Psychometric assessment of the Arabic version of the Internalized Stigma of Mental Illness (ISMI) measure in a refugee population. *Transcultural Psychiatry* 52, 636–658. doi: 10.1177/1363461515569755.
- Kobau R, DiIorio C, Chapman D, Delvecchi P** (2010). SAMHSA/CDC Mental Illness Stigma Panel Members. Attitudes about mental illness and its treatment: validation of a generic scale for public health surveillance of mental illness associated stigma. *Community Mental Health Journal* 46, 164–176. doi: 10.1007/s10597-009-9191-x.
- Komiya N, Good GE, Sherrod NB** (2000). Emotional openness as a predictor of college students' attitudes toward seeking psychological help. *Journal of Counseling Psychology* 47, 138–143. doi: 10.1037//0022-0167.47.1.138.
- Kutcher S, Bagnell A, Wei Y** (2015a). Mental health literacy in secondary schools: a Canadian approach. *Child and Adolescent Psychiatric Clinics* 24, 233–244.
- Kutcher S, Wei Y, Morgan C** (2015b). Successful application of a Canadian mental health curriculum resource by usual classroom teachers in significantly and sustainably improving student mental health literacy. *Canadian Journal of Psychiatry* 60, 580–586.
- Kutcher S, Wei Y, Coniglio C** (2016). Mental health literacy: past, present, and future. *Canadian Journal of Psychiatry* 61, 154–158.
- Lam DCK, Salkovskis PM, Warwick HMC** (2005). An experimental investigation of the impact of biological versus psychological explanations of the cause of "mental illness". *Journal of Mental Health* 14, 453–464. doi: 10.1080/09638230500270842.
- Lien Y, Kao Y, Liu Y, Chang H, Tzeng N, Lu C, Loh CH** (2014). Internalized stigma and stigma resistance among patients with mental illness in Han Chinese population. *Psychiatric Quarterly* 86, 181–197. doi: 10.1007/s11226-014-9315-5.
- Link BG** (1987). Understanding labeling effects in the area of mental disorders: an assessment of the effects of expectations of rejection. *American Sociological Review* 52, 96–112.
- Link BG, Cullen FT, Frank J, Wozniak JF** (1987). The social rejection of former mental patients: understanding why labels matter. *American Journal of Sociology* 92, 1461–1500.
- Link BG, Cullen FT, Struening E, Shroud PE, Dohrenwend BP** (1989). A modified labeling theory approach to mental disorders: an empirical assessment. *American Sociological Review* 54, 400–423.
- Link BG, Mirotznik J, Cullen FT** (1991). The effectiveness of stigma coping orientations: can negative consequences of mental illness labeling be avoided? *Journal of Health and Social Behavior* 32, 302–320.
- Link BG, Struening EL, Rahav M, Phelan JC, Nuttbrock L** (1997). On stigma and its consequences: evidence from a longitudinal study of men with dual diagnoses of mental illness and substance abuse. *Journal of Health Social Behavior* 38, 177–190.
- Link BG, Yang LH, Phelan JC, Collins PY** (2004). Measuring mental illness stigma. *Schizophrenia Bulletin* 30, 511–541.
- Luca Pingani L, Forghieri M, Ferrari S, Ben-Zeev D, Artoni P, Mazzi F, Palmieri G, Rigatelli M, Corrigan PW** (2012). Stigma and discrimination toward mental illness: translation and validation of the Italian version of the attribution questionnaire-27 (AQ-27-I). *Social Psychiatry and Psychiatric Epidemiology* 47, 993–999. doi: 10.1007/s00127-011-0407-3.

- Luty J, Fekadu D, Umoh O, Gallagher J (2006). Validation of a short instrument to measure stigmatized attitudes towards mental illness. *Psychiatric Bulletin* **30**, 257–260. doi: 10.1192/pb.30.7.257.
- Maccoby EE and Maccoby N (1954). The interview: a tool of social science. In *Handbook of Social Psychology*, Vol. I. (ed. G. Lindzey), pp. 449–487. Addison-Wesley: Cambridge, MA.
- Madianos MG, Madianou D, Vlachonikolis J, Stefanis CN (1987). Attitudes towards mental illness in the Athens area: implications for community mental health intervention. *Acta Psychiatrica Scandinavica* **75**, 158–165.
- Madianos M, Economou M, Peppou LE, Kallergis G, Rogakou E, Alevizopoulos G (2012). Measuring public attitudes to severe mental illness in Greece: development of a new scale. *European Journal of Psychiatry* **26**, 55–67.
- Magliano L, Marasco C, Guarneri M, Malangone C, Lacrimini G, Zanus P, et al. (1999). A new questionnaire assessing the opinions of the relatives of patients with schizophrenia on the causes and social consequences of the disorder: reliability and validity. *European Psychiatry* **14**, 71–75.
- Mak WWS, Cheung RYM (2008). Affiliate stigma among caregivers of people with intellectual disability or mental illness. *Journal of Applied Research in Intellectual Disabilities* **21**, 532–545. doi: 10.1111/j.1468-3148.2008.00426.x.
- Mak WWS, Cheung RYM (2010). Self-stigma among concealable minorities in Hong Kong: conceptualization and unified measurement. *American Journal of Orthopsychiatry* **80**, 267–281. doi: 10.1111/j.1939-0025.2010.01030.x.
- Mansouri L, Dowell DA (1989). Perceptions of stigma among the long-term mentally ill. *Psychosocial Rehabilitation Journal* **13**, 79–91.
- McKeague L, Hennessy E, O'Driscoll C, Heary C (2015). Peer Mental Health Stigmatization Scale: psychometric properties of a questionnaire for children and adolescents. *Child and Adolescent Mental Health* **20**, 163–170. doi: 10.1111/camh.12088.
- McLuckie A, Kutcher S, Wei Y, Weaver C (2014). Sustained improvements in students' mental health literacy with use of a mental health curriculum in Canadian schools. *BMC Psychiatry* **14**, 1694. doi: 10.1186/s12888-014-0379-4.
- Michaels PJ, Corrigan PW (2013). Measuring mental illness stigma with diminished social desirability effects. *Journal of Mental Health* **22**, 218–226. doi: 10.3109/09638237.2012.734652.
- Milin R, Kutcher S, Lewis SP, Walker S, Wei Y, Ferrill N, Armstrong M (2016). Impact of a mental health curriculum on knowledge and stigma among high school students: a randomized controlled trial. *Journal of American Academy of Child and Adolescent Psychiatry* **55**, 383–391.
- Minnebo J, Acker AV (2004). Does television influence adolescents' perceptions of and attitudes toward people with mental illness? *Journal of Community Psychology* **32**, 257–275. doi: 10.1002/jcop.20001.
- Modgill G, Knaak S, Kassam A, Szeto A (2014). Opening minds stigma scale for health care providers (OMS-HC): examination of psychometric properties and responsiveness. *BMC Psychiatry* **14**, 120. doi: 10.1186/1471-244X-14-120.
- Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Journal of Clinical Epidemiology* **62**, 1006–1012.
- Morris R, Scott PA, Cocoman A, Chambers M, Guise V, Välimäki M, Clinton G (2011). Is the Community Attitudes towards the Mentally Ill scale valid for use in the investigation of European nurses' attitudes towards the mentally ill? A confirmatory factor analytic approach. *Journal of Advanced Nursing* **68**, 460–470. doi: 10.1111/j.1365-2648.2011.05739.x.
- Morrison JK, Becker BE (1975). Seminar-induced change in a community psychiatric team's reported attitudes toward "mental illness". *Journal of Community Psychology* **3**, 281–284.
- Moses T (2009). Stigma and self-concept among adolescents receiving mental health treatment. *American Journal of Orthopsychiatry* **79**, 261–274. doi: 10.1037/a0015696.
- Nevid JS, Morrison J (1980). Attitudes toward mental illness: the construction of the libertarian mental health ideology scale. *Journal of Humanistic Psychology* **20**, 71–85. doi: 10.1177/002216788002000207.
- Ng P, Chan K (2000). Sex differences in opinion towards mental illness of secondary school students in Hong Kong. *International Journal of Social Psychiatry* **46**, 79–88. doi: 10.1177/002076400004600201.
- Patel V, Flisher AJ, Hetrick S, McGorry P (2007). Mental health of young people: a global public-health challenge. *Lancet* **369**, 1302–1313.
- Penn DL, Guynan K, Daily T, Spaulding WD, Garbin CP, Sullivan M (1994). Dispelling the stigma of schizophrenia: what sort of information is best? *Schizophrenia Bulletin* **20**, 567–577.
- Pinto MD, Hickman R, Logsdon MC, Burant C (2012). Psychometric evaluation of the revised attribution questionnaire (r-AQ) to measure mental illness stigma in adolescents. *Journal of Nursing Measurement* **20**, 47–58. doi: 10.1891/1061-3749.20.1.47.
- RefWorks-COS PL, ProQuest LLC (2001). *RefWorks*, 2nd edn. ProQuest LLC: Ann Arbor, MI.
- Ritsher JB, Phelan JC (2004). Internalized stigma predicts erosion of morale among psychiatric outpatients. *Psychiatry Research* **129**, 257–265.
- Ritsher JB, Otilingama PG, Grajales M (2003). Internalized stigma of mental illness: psychometric properties of a new measure. *Psychiatry Research* **121**, 31–49. doi: 10.1016/j.psychres.2003.08.008.
- Serra M, Lai A, Buizza C, Pioli R, Preti A, Masala C, Petretto DR (2013). Beliefs and attitudes among Italian high school students toward people with severe mental disorders. *Journal of Nervous and Mental Disease* **201**, 311–318.
- Sevigny R, Yang W, Zhang P, Marleau JD, Yang Z, Lin S, Li G, Xu D, Wang Y, Wang H (1999). Attitudes toward the mentally ill in a sample of professionals working in a psychiatric hospital in Beijing. *International Journal of Social Psychiatry* **45**, 41. doi: 10.1177/002076409904500106.
- Scheerder G, De Coster I, Van Audenhove C (2009). Community pharmacists' attitude toward depression: a pilot study. *Research in Social and Administrative Pharmacy* **5**, 242–252.
- Schneider J, Beeley C, Repper J (2011). Campaign appears to influence subjective experience of stigma. *Journal of Mental Health* **20**, 89–97. doi: 10.3109/09638237.2010.537403.

- Sibitz I, Amering M, Unger A, Seyringer ME, Bachmann A, Schrank B, Benesch T, Schulze B, Woppmann A** (2011a). The impact of the social network, stigma and empowerment on the quality of life in patients with schizophrenia. *European Psychiatry* **26**, 28–33.
- Sibitz I, Unger A, Woppmann A, Zidek T, Amering M** (2011b). Stigma resistance in patients with schizophrenia. *Schizophrenia Bulletin* **37**, 316–323.
- Sorsdahl KR, Kakuma R, Wilson Z, Stein DJ** (2012). The internalized stigma experienced by members of a mental health advocacy group in South Africa. *International Journal of Social Psychiatry* **58**, 55. doi: 10.1177/0020764010387058.
- Streiner DL, Norman GR** (2008). *Health Measurement Scales: a Practical Guide to their Development and Use*, 4th edn. Oxford University Press: New York.
- Struening EL, Cohen J** (1963). Factorial invariance and other psychometric characteristics of five opinions about mental illness factors. *Educational and Psychological Measurement* **23**, 289–298. doi: 10.1177/001316446302300206.
- Struening EL, Perlick DA, Link BG, Hellman FH, Herman D, Sirey JA** (2001). The extent to which caregivers believe most people devalue consumers and their families. *Psychiatric Services* **52**, 1633–1638.
- Stuart H, Milev R, Koller M** (2005). The inventory of stigmatizing experiences: its development and reliability. *World Psychiatry* **4**: S1, 35–39.
- Svensson B, Markström U, Bejerholm U, Björkman T, Brunt D, Eklund M, Hansson L, Leufstadius C, Gyllensten AL, Sandlund M, Ostman M** (2011). Test – retest reliability of two instruments for measuring public attitudes towards persons with mental illness. *BMC Psychiatry* **11**, 11. <http://www.biomedcentral.com/1471-244X/11/11>. Accessed 14 August 2015.
- Swami V, Furnham A** (2011). Preliminary examination of the psychometric properties of the Psychiatric Scepticism Scale. *Scandinavian Journal of Psychology* **52**, 399–403. doi: 10.1111/j.1467-9450.2011.00881.x.
- Świtaj P, Paweł Grygiel P, Jacek Wciórka J, Humenny G, Anczewska M** (2013). The stigma of subscale of the Consumer Experiences of Stigma Questionnaire (CESQ): a psychometric evaluation in Polish psychiatric patients. *Comprehensive Psychiatry* **54**, 713–719. doi: 10.1016/j.comppsy.2013.03.001.
- Taylor SM, Dear MJ** (1981). Scaling community attitudes toward the mentally ill. *Schizophrenia Bulletin* **7**, 226–240.
- Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM, de Vet HC** (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology* **60**, 34–42.
- Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM** (2012). Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Quality of Life Research* **21**, 651–657. doi: 10.1007/s11136-011-9960-1.
- Thornicroft G** (2006). *Shunned: Discrimination Against People with Mental Illness*. OUP Oxford: London.
- Thornicroft G, Brohan E, Rose D, Sartorius N, Leese M, For the INDIGO Study Group** (2009). Global pattern of experienced and anticipated discrimination against people with schizophrenia: a cross-sectional survey. *Lancet* **373**, 408–415.
- Thornicroft G, Mehta N, Clement S, Evans-Lacko S, Doherty M, Rose D, Koschorke M, Shidhaye R, O'Reilly C, Henderson C** (2016). Evidence for effective interventions to reduce mental-health-related stigma and discrimination. *Lancet* **387**, 1123–1132. doi: 10.1016/S0140-6736(15)00298-6.
- Uijen AA, Heinst CW, Schellevis FG, van den Bosch WJ, van de Laar FA, Terwee CB, Schers HJ** (2012). Measurement properties of questionnaires measuring continuity of care: a systematic review. *PLoS ONE* **7**, e42256.
- Vega WA, Rodriguez MA, Ang A** (2010). Addressing stigma of depression in Latino primary care patients. *General Hospital Psychiatry* **32**, 182–191. doi: 10.1016/j.genhosppsy.2009.10.008.
- Vogel DL, Wade NG, Haake S** (2006). Measuring the self-stigma associated with seeking psychological help. *Journal of Counseling Psychology* **53**, 325–337. doi: 10.1037/0022-0167.53.3.325.
- Vogel DL, Wade NG, Aschman PL** (2009). Measuring perceptions of stigmatization by others for seeking psychological help: reliability and validity of a new stigma scale with college students. *Journal of Counseling Psychology* **56**, 301–308. doi: 10.1037/a0014903.
- Vogt D, Di Leone BAL, Wang JM, Sayer NA, Pineles SL** (2014). Endorsed and anticipated stigma inventory (EASI): a tool for assessing beliefs about mental illness and mental health treatment among military personnel and veterans. *Psychological Services* **11**, 105–113. doi: 10.1037/a0032780.
- Watson AC, Miller FE, Lyons JS** (2005). Adolescent attitudes toward serious mental illness. *Journal of Nervous and Mental Disease* **193**, 769–772. doi: 10.1097/01.nmd.0000185885.04349.99.
- Wei Y, McGrath P, Hayden J, Kutcher S** (2015). Mental health literacy measures evaluating knowledge, attitudes and help-seeking: a scoping review. *BMC Psychiatry* **15**, 291. doi: 10.1186/s12888-015-0681-9.
- Wolff G, Pathare S, Graig T, Leff J** (1996). Community attitudes to mental illness. *British Journal of Psychiatry* **168**, 183–190.
- World Health Organization** (2011). Global burden of mental disorders and the need for a comprehensive, coordinated response from health and social sectors at the country level. [http://apps.who.int/gb/ebwha/pdf\\_files/EB130/B130\\_9-en.pdf](http://apps.who.int/gb/ebwha/pdf_files/EB130/B130_9-en.pdf). Accessed 10 June 2016.
- Wu TH, Chang CC, Chen CY, Wang JD, Lin CY** (2015). Further psychometric evaluation of the self-stigma scale – short: measurement invariance across mental illness and gender. *PLoS ONE* **10**, e0117592. doi: 10.1371/journal.pone.0117592.
- Yamaguchi S, Koike S, Watanabe K, Ando S** (2014). Development of a Japanese version of the reported and intended behavior scale: reliability and validity. *Psychiatry Clinical Neuroscience* **68**, 448–455. doi: 10.1111/pcn.12151.
- Zisman-Ilani Y, Levy-Frank-Levy I, Hasson-Ohayon I, Kravetz S, Mashiach-Eizenberg M, Roe D** (2013). Measuring the internalized stigma of parents of persons with a serious mental illness. *Journal of Nervous and Mental Disease* **1**, 183–187.

## Appendix 1: Search strategies in PubMed

Concept 1 AND Concept 2 AND Concept 3 AND Concept 4 Key Mental health disorders and mental health	3 aspects of MHL	Assessment tool	Study type
OR "Mental Disorders"[Mesh: noexp] OR "mental health"[Mesh: noexp] "Substance-related disorders"[Mesh] OR substance use disorder*[tiab] OR "substance abuse"[tiab] OR "substance misuse"[tiab] OR "substance dependence"[tiab]	"health education"[tiab] "health education"[Mesh]	assessment*[tiab] evaluat*[tiab]	Reliability[tiab] effective*[tiab]
OR anxiety disorder*[tiab] OR "anxiety disorders"[Mesh] OR "generalized anxiety disorder"[tiab] OR "separation anxiety disorder"[tiab] OR "social phobia"[tiab] OR "specific phobia"[tiab] OR "panic disorder"[tiab] OR "posttraumatic stress disorder"[tiab]	"mental health literacy"[tiab]	measur*[tiab]	efficac*[tiab]
OR disruptive behavior disorder*[tiab] OR "attention deficit and disruptive behavior disorders"[Mesh] OR "conduct disorder"[tiab] OR "oppositional defiant disorder"[tiab]	"health knowledge"[tiab]	test*[tiab]	"program evaluation"[Mesh] OR "program evaluation"[tiab]
OR "unipolar depression"[tiab] OR "major depressive disorder"[tiab] OR depression[tiab] OR "depressive disorder"[Mesh] OR "depression"[Mesh]	"health curriculum"[tiab]	scale*[tiab]	Validity[tiab]
OR "attention deficit hyperactivity disorder"[tiab] OR ADHD[tiab]	"mental health awareness"[tiab] awareness[Mesh]	assessment tool*[tiab] psychometrics[Mesh] OR psychometrics[tiab]	
OR	"attitude to health"[Mesh]	questionnaires[Mesh] OR questionnaire*[tiab]	
OR		survey*[tiab]	
OR	stigma[tiab]		
OR	discrimination[tiab]		
OR	"help seeking behavior"[tiab] OR "seeking help"[tiab]		

Appendix 2: Quality criteria of measurement properties (Terwee *et al.* 2007)

Property	Quality criteria	Rating
Reliability		
Internal consistency	(Sub)scale unidimensional AND Cronbach's alpha(s) $\geq 0.70$	+
	Dimensionality not known OR Cronbach's alpha not determined	?
	(Sub)scale not unidimensional OR Cronbach's alpha(s), 0.70	-
	Positive rating (+) in one subgroup, however negative rating (-) or unknown (?) in another subgroup in the same study	+/-
Reliability	ICC/weighted Kappa $\geq 0.70$ OR Pearson's $r \geq 0.80$	+
	Neither ICC/weighted Kappa, nor Pearson's $r$ determined	?
	ICC/weighted Kappa $\leq 0.70$ OR Pearson's $r \leq 0.80$	-
	Positive rating (+) in one subgroup, however negative rating (-) or unknown (?) in another subgroup in the same study	+/-
Measurement error	MIC > SDC OR MIC outside the LOA	+
	MIC not defined	?
	MIC $\leq$ SDC OR MIC equals or inside LOA	-
	Positive rating (+) in one subgroup, however negative rating (-) or unknown (?) in another subgroup in the same study	+/-
Validity		
Content validity	The target population considers all items in the questionnaire to be relevant AND considers the questionnaire to be complete	+
	No target population involvement	?
	The target population considers items in the questionnaire to be irrelevant OR considers the questionnaire to be incomplete	-
	Positive rating (+) in one subgroup, however negative rating (-) or unknown (?) in another subgroup in the same study	+/-
Structural validity	Factors should explain at least 50% of the variance	+
	Explained variance not mentioned	?
	Factors explain <50% of the variance	-
	Positive rating (+) in one subgroup, however negative rating (-) or unknown (?) in another subgroup in the same study	+/-
Hypothesis testing (construct validity)	Correlation with an instrument measuring the same construct $\geq 0.50$ OR at least 75% of the results are in accordance with the hypotheses AND correlation with related constructs is higher than with unrelated constructs	+
	Solely correlations determined with unrelated constructs	?
	Correlation with an instrument measuring the same construct <0.50 OR <75% of the results are in accordance with the hypotheses OR correlation with related constructs is lower than with unrelated constructs	-
	Positive rating (+) in one subgroup, however negative rating (-) or unknown (?) in another subgroup in the same study	+/-
Criterion validity	Correlations with the gold standard is $\geq 0.70$	+
	Correlations with the gold standard is unknown	?
	Correlations with the gold standard is <0.70	-
	Positive rating (+) in one subgroup, however negative rating (-) or unknown (?) in another subgroup in the same study	+/-
Responsiveness		
Responsiveness	(Correlation with an instrument measuring the same construct $\geq 0.50$ OR at least 75% of the results are in accordance with the hypotheses OR AUC $\geq 0.70$ ) AND correlation with related constructs is higher than with unrelated constructs	+
	Solely correlations determined with unrelated constructs	?
	Correlation with an instrument measuring the same construct <0.50 OR <75% of the results are in accordance with the hypotheses OR AUC <0.70 OR correlation with related constructs is lower than with unrelated constructs	-
	Positive rating (+) in one subgroup, however negative rating (-) or unknown (?) in another subgroup in the same study	+/-

**Appendix 3: Levels of evidence for the overall quality of the measurement property (Uijen et al. 2012; Furlan et al. 2015)**

Level	Rating	Criteria
Strong	+++ or ---	Consistent findings in multiple studies of good methodological quality OR in one study of excellent methodological quality
Moderate	++ or --	Consistent findings in multiple studies of fair methodological quality OR in one study of good methodological quality
Limited	+ or -	One study of fair methodological quality
Conflicting	+/-	Conflicting findings
Unknown	x	Studies of poor methodological quality or studies with indeterminate rating of the measurement property