Original Research

The Psychometric Assessment of Internalized Stigma Instruments: A Systematic Review

Sharon Anne Maria Stevelink BSc, MSc[1,*], Ingeborg Charlotte Wu BSc, MSc[1], Carlijn Gabriëlle Nicolette Voorend Voorend BSc, MSc[2], Wim Herman van Brakel MD, MSc, PhD[1,3]

[1]Athena Institute for Research on Innovation and Communication in Health and Life Sciences, Faculty of Earth and Life Sciences, VU University Amsterdam, Netherlands
[2]Netherlands Leprosy Relief, Amsterdam, Netherlands
[3]Royal Tropical Institute, Leprosy Unit, Amsterdam, Netherlands

Abstract

Objective: To rate the psychometric properties of instruments to measure internalized (or self-) stigma in health conditions where stigma plays a major role.

Method: We conducted a systematic literature review by searching relevant databases and by reviewing the bibliographies of relevant papers. Quantitative studies were included if the items used, or a sample of the instrument, was included in the paper and if the studies focussed on the initial development or validation of the instrument. Health conditions included were HIV/AIDS, mental health, leprosy, asthma, epilepsy, cancer, obesity, and tuberculosis. Psychometric properties of the included studies were assessed using the quality criteria proposed by Terwee *et al.* and the COSMIN consortium: content validity, internal consistency, construct validity, criterion validity, reproducibility, responsiveness, floor and ceiling effects, and interpretability.

Results: Thirty-three papers were included of which 21 were identified as actual instrument development studies. Only two instruments received three positive quality ratings, 12 received at least three indeterminate ratings, especially for the internal consistency and construct validity. At least one negative rating was given to five instruments. Content and construct validity as well as internal consistency were most often assessed, whereas agreement and responsiveness received least attention.

Conclusions: We rated the psychometric properties of available instruments to measure internalized stigma using standard quality criteria. Only the Child Attitude Towards Illness Scale and the Internalized Stigma of Mental Illness received three positive ratings indicating that the majority of the instruments need further testing.

Implications: The need was identified for a simplified testing protocol to design an instrument development study, to assess certain psychometric properties, and to specify the preferred statistical methods for testing these. In addition, researchers should be aware that re-validation of instruments is necessary before they are used in cultures and study populations other than those for which they were developed.

Keywords: health-related stigma, measurement, psychometric properties, reliability, scale, validity

Introduction

Health-related stigma describes a social concept that has an enormous impact on the lives of many people that suffer from a certain health condition (Van Brakel, 2006). Several components of stigma can be distinguished. According to Rensen, Bandyopadhyay, Gopal & Van Brakel (2010), stigma can be categorized from the perspective of the non-affected person into perceived and enacted stigma, and from the perspective of the affected person into internalized, perceived, and experienced stigma (Rensen *et al.*, 2010). These different aspects are all interrelated and may have an impact on the self-efficacy of the affected person, his or her

*Corresponding author: Sharon Stevelink
E-mail: s.a.m.stevelink@vu.nl; sharonstevelink@hotmail.com

participation in the community, personal well-being, and self-esteem (Rensen *et al.*, 2010).

Various studies have reported substantial levels of internalized stigma across health conditions (Arole, Premkumar, Arole, Maury & Saunderson, 2002; Dinos, Stevens, Serfaty, Weich & King, 2004; Adewuya, Owoeye, Erinfolami & Ola, 2010; Rensen *et al.*, 2010; Sorsdahl, Mall, Stein & Joska, 2011; Stevelink, Van Brakel & Augustine, 2011). Affected persons may feel disappointed for contracting the particular disease or condition, and feel embarrassed, guilty, and inferior compared to others (Arole *et al.*, 2002; Dinos *et al.*, 2004; Sorsdahl *et al.*, 2011; Stevelink *et al.*, 2011). Internalized stigma is defined as a 'subjective process, embedded within a socio-cultural context, which may be characterized by negative feelings (about self), maladaptive behavior, identity transformation or stereotype endorsement resulting from an individual's experiences, perceptions, or anticipation of negative social reaction on the basis of their health condition' (Livingston & Boyd, 2010). Feelings may involve loss of self-esteem, isolation, fear, etc., (Corrigan, 1998; Van Brakel, 2006). A modified conceptualization of this concept can be found in Fig. 1.

Several instruments have been developed to measure internalized stigma for use in different health areas such as HIV/AIDS, leprosy, mental health conditions, epilepsy, obesity and cancer. The process of finding the most suitable instrument for a particular study is complex and time-consuming due to the large body of research done. This literature review provides a comprehensive overview and psychometric assessment of the best validated instruments to measure internalized stigma.

Methods

References were identified through an extensive and systematic search in *Pubmed (Medline), Web of Science, PsycINFO* and *WorldCat.* We used a generic syntax consisting of main key words present in the title, abstract or main text. This syntax was a variation of the following: <stigma AND (measure* OR assess* OR instrument OR question* OR scale)> and <("self stigma" OR "internal* stigma" OR "personal stigma") AND (measure* OR assess* OR instrument OR question* OR scale)>. Additional studies were identified by scanning relevant bibliographies and personal communications with experts in the field. The last search was done on 8 June 2011.

One author reviewed the titles and abstracts for relevance. After this first selection, full-text articles were reviewed and checked to determine if they met the inclusion criteria. This review was limited to quantitative English language studies, and only the studies describing the actual development of the instrument(s) were included in the literature review. Whenever available, studies were included if they provided additional validation of the instrument such as in a different health condition or language. Furthermore, instruments were

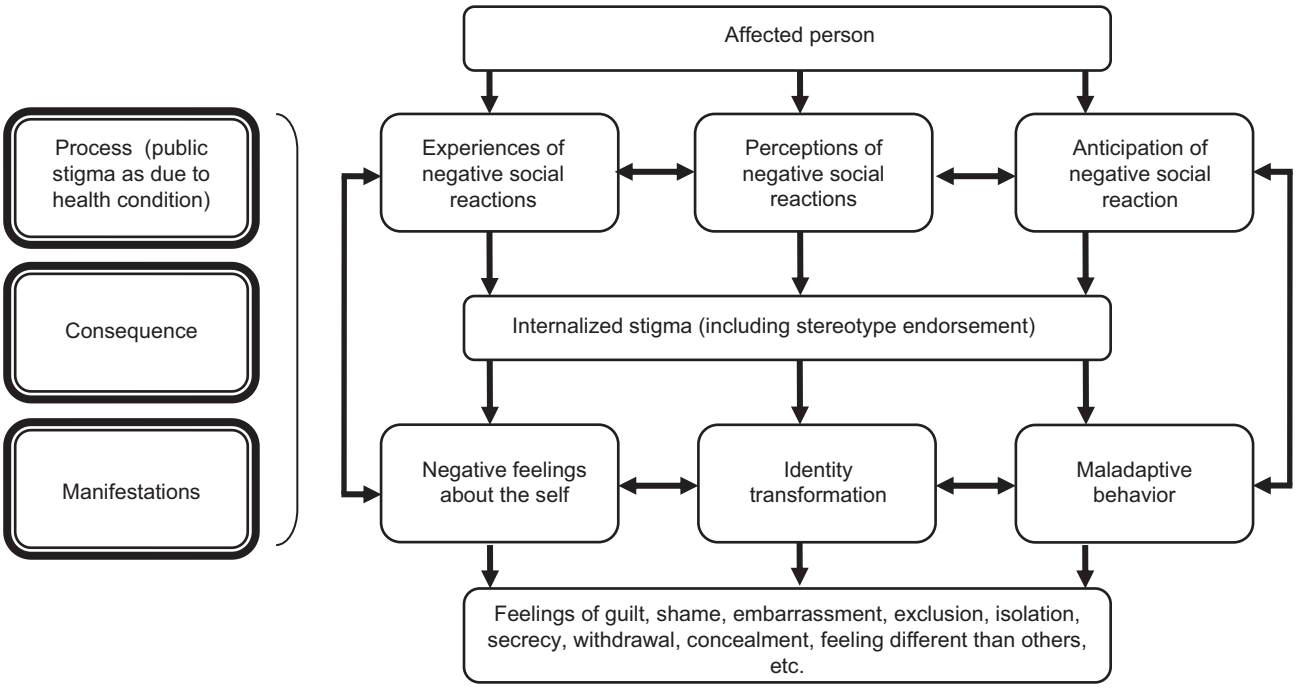

Fig. 1. Adapted framework internalized stigma (based on Livingston & Boyd, 2010).

included only if they specifically aimed to measure internalized stigma with a (sub) scale or a related construct.

To provide an adequate overview of the best validated instruments, a quality criteria framework was used. This framework, developed by Terwee *et al.* (2007), provides indications for what constitutes 'good' psychometric properties. International consensus about the exact terminology, taxonomy, and definitions of these properties was reached by a consortium of experts (Mokkink *et al.*, 2010a,b). We incorporated their findings within the quality criteria framework applied this in this review. The following psychometric properties were assessed: content validity, internal consistency, criterion validity, construct validity, reproducibility, responsiveness, floor and ceiling effects, and interpretability (Terwee *et al.*, 2007; Mokkink *et al.*, 2010a). The framework states exactly which statistical methods to use for addressing the psychometric properties and the criteria for rating these properties as 'positive', 'indeterminate', 'negative' or 'no information available'. Two raters assessed each instrument independently after which identified discrepancies were discussed. If no consensus was reached a third rater was consulted. An extended description of these quality criteria can be found in Table 1.

Results

A total of 1,036 articles were identified from the first search. After a selection on title and abstract, the full text of 55 articles was assessed. Finally, 33 articles, including 21 different instruments, were included. An overview of the selected instruments can be found in Table 2 and a summary of the ratings assigned in Table 3. This summarized rating provides an indication of which properties have been addressed in the different studies that made use of the instrument. The best results found for a particular property are included in this table. Individual ratings assigned per study are available from the authors upon request.

HIV/AIDS

The 13-item HIV Stigma Scale was developed by Sowell *et al.* (1997) and further validated by Emlet (2005) for use in older adults living with HIV/AIDS. Three subscales can be distinguished: distancing, blaming, and discrimination (Emlet, 2005). During the initial development study, content validity was addressed by the active involvement of the target population during focus group discussions concerning item generation, selection, and ease of item understanding (Sowell *et al.*, 1997). During the further validation study, the factor structure of the scale was defined and Cronbach's alphas calculated. However, this was rated as indeterminate, because the sample size for the factor analysis was borderline insufficient (13 items, n=88, instead of n=91 (7 times the number of items). In addition, two alphas of the subscales were below the quality threshold (α=0.70), respectively 0.60, 0.76 and 0.62 (Emlet, 2005). Construct validity was addressed and the correlations were as expected, however, the exact magnitude of the hypothesized correlations was not defined, resulting in an indeterminate rating (Emlet, 2005).

Fife & Wright (2000) developed the Social Impact Scale (SIS) and tested this scale simultaneously in samples of persons with HIV/AIDS and persons with cancer. The 24-item scale comprises four subscales, identified during principal component extraction (n=206): social rejection, internalized shame, social isolation, and financial insecurity. All Cronbach's alphas were sufficient (ranging from 0.85 to 0.90), which resulted in a positive rating for internal consistency. In addition, content validity was rated positively. Short and simple items were generated and selected by an expert panel, after which these were pilot-tested with the help of members of the target populations (Fife & Wright, 2000).

A Chinese version of this scale was further validated for use with persons affected by depression, schizophrenia and HIV/AIDS (Pan *et al.*, 2007). Content validity was rated indeterminate, because members of the target populations were not involved during the translation process. No factor analysis was applied and, according to the results of a Rasch analysis, the SIS was found to be an unidimensional scale (Pan *et al.*, 2007). In 2008, the SIS was further validated for use in persons with Alzheimer's dementia and Parkinson's disease (Burgener & Berger, 2008). This study did not apply factor analysis and some Cronbach's alphas were below the threshold of 0.70. Construct validity was assessed with the help of instruments measuring self-esteem, depression, and personal control. However, no adequate *a priori* hypotheses were formulated (e.g., exact magnitudes of the expected correlations were not hypothesized), which resulted in an indeterminate rating (Burgener & Berger, 2008).

A widely applied scale developed to measure internalized, perceived, and enacted stigma, is the HIV Stigma Scale developed by Berger *et al.* (2001). Initially, the scale consisted of 40 items, divided into four subscales: personalized (enacted) stigma, disclosure concerns, negative self-image, and concern with public attitudes (Berger *et al.*, 2001). Several validation studies developed an abbreviated version of the scale,

Table 1: Adapted Overview Quality Criteria Proposed by Terwee *et al.* (2007)

Property	Definition	Quality Criteria[a,b]
1. Content Validity	The extent to which the domain of interest is comprehensively sampled by the items in the questionnaire	+ A clear description is provided of the measurement aim, the target population, the concepts that are being measured, and the item selection AND target population and (investigators OR experts) were involved in item selection; **?** A clear description of above-mentioned aspects is lacking or only target population involved OR doubtful design or method; **-** No target population involvement; **0** No information found on target population involvement.
2. Internal Consistency	The extent to which items in a (sub)scale are intercorrelated, thus measuring the same construct	+ Factor analyses performed on adequate sample size (7 x # items and ≥100) AND Cronbach's alpha(s) calculated per dimensions AND Cronbach's alpha(s) between 0.70 and 0.95; **?** No factor analysis OR doubtful design or method; **-** Cronbach's alpha(s) <0.70 or >0.95, despite adequate design and method; **0** No information found on internal consistency.
3. Criterion Validity	The extent to which scores on a particular questionnaire relate to a golden standard	+ Convincing arguments that gold standard is 'gold' AND correlation with gold standard ≥0.70; **?** No convincing arguments that gold standard is 'gold' OR doubtful design or method; **-** Correlation with gold standard <0.70, despite adequate design and method; **0** No information found on criterion validity.
4. Construct Validity	The extent to which scores on a particular questionnaire relate to other measures in a manner that is consistent with theoretically derived hypotheses concerning the concepts that are being measured	+ Specific hypotheses were formulated *a priori* AND include direction AND exact magnitude of the correlation AND at least 75% of the results are in accordance with these hypotheses; **?** Doubtful design or method (e.g., no hypotheses); **-** Less than 75% of hypotheses were confirmed, despite adequate design and methods; **0** No information found on construct validity.
5. Reproducibility		
5.1 Agreement	The extent to which the scores on repeated measures are close to each other (absolute measurement error)	+ MIC<SDC or MIC outside LOA OR convincing arguments that agreement is acceptable; **?** Doubtful design or method OR (MIC not defined AND no convincing arguments that agreement is acceptable); **-** MIC ≥SDC OR MIC equals inside LOA, despite adequate design and method; **0** No information found on agreement.
5.2 Reliability	The extent to which patients can be distinguished from each other, despite measurement errors (relative measurement error)	+ ICC or weighted kappa ≥0.70; **?** Doubtful design or method (e.g., time interval not mentioned); **-** ICC or weighted kappa<0.70, despite adequate design and method; **0** No information found on reliability.
6. Responsiveness	The ability of a questionnaire to detect clinically important changes over time	+ SDC or SDC < MIC or MIC outside LOA or RR>1.96 OR AUC ≥0.70; **?** Doubtful design or method; **-** SDC or SDC ≥MIC or MIC equals or inside LOA OR RR ≤1.96 or AUC <0.70, despite adequate design and methods; **0** No information found on responsiveness.

(Continued on next page)

Table 1: *(Continued)*

Property	Definition	Quality Criteria[a,b]
7. Floor and Ceiling Effects	The number of respondents who achieved the lowest or highest possible score	+ ≥15% of the respondents achieved the highest or lowest possible scores; ? Doubtful design or method; - 15% of the respondents achieved the highest or lowest possible scores, despite adequate design and methods; 0 No information found on floor and ceiling effects.
8. Interpretability	The degree to which one can assign qualitative meaning to quantitative scores	+ Mean and SD scores presented of at least four relevant subgroups of patients and MIC defined; ? Doubtful design or method OR less than four subgroups OR no MIC defined; 0 No information found on interpretation.

Notes: MIC = minimal important change; SDC = smallest detectable change; LOA = limits of agreement; ICC = intraclass correlation; SD, standard deviation.
[a]+ = positive rating; ? = indeterminate rating; - = negative rating; 0 = no information available. [b]Doubtful design or method = lacking of clear description of the design or methods of the study, sample size smaller than 50 subjects (should be at least 50 in every (subgroup) analysis, or any important methodological weakness in the design or execution of the study.

resulting in 32-, 21-, 17- and 10-item versions (Bunn *et al.*, 2007; Wright *et al.*, 2007; Franke *et al.*, 2010; Jimenez *et al.*, 2010). Content validity was well established. Items were generated with help of an extensive literature review and expert consultation, after which items were pre-tested for relevancy and readability with the help of members of the target population (Berger *et al.*, 2001). In general, Cronbach's alphas were found to be acceptable (Bunn *et al.*, 2007; Wright *et al.*, 2007; Franke *et al.*, 2010; Jimenez *et al.*, 2010). However, in the studies that applied factor analysis, the sample size was not sufficient, which resulted in an indeterminate rating (Bunn *et al.*, 2007; Wright *et al.*, 2007; Franke *et al.*, 2010; Jimenez *et al.*, 2010). Construct validity was assessed in the majority of the studies by calculating hypothesized correlations with measures for depression, social support, etc., In general, these hypotheses conformed to expectations (Bunn *et al.*, 2007; Wright *et al.*, 2007; Franke *et al.*, 2010; Jimenez *et al.*, 2010). Despite this, an indeterminate rating was provided for construct validity because the magnitudes of the expected correlations were not defined. Responsiveness and reliability were rated as indeterminate also, because no adequate statistical methods were used to assess these properties (Franke *et al.*, 2010; Jimenez *et al.*, 2010).

The 33-item HIV/AIDS Stigma Instrument – PLWA (HASI-P) has several subscales, including a subscale for internalized stigma (Holzemer *et al.*, 2007). The items were developed with help of focus group discussions with persons living with HIV/AIDS. Factor analysis was applied on an insufficient sample size (90 items, n=217 and 72 items, n=1477), resulting in an indeterminate rating. Cronbach's alphas were sufficient (0.76–0.91). Construct validity was tested and hypotheses formulated *a priori*, however, the exact formulation of the magnitude and the expected direction was lacking, resulting in an indeterminate rating (Holzemer *et al.*, 2007).

Sayles *et al.* (2008) specifically developed an instrument to measure internalized stigma. After several focus groups and cognitive interviews with a diverse group of people living with HIV/AIDS, 52 items were retained in the first version of the scale (Sayles *et al.*, 2008). After the application of exploratory factor analysis (52 items, n=202), the 28-item final version of the Internalized HIV Stigma Measure was formulated, consisting of four subscales: stereotypes, disclosure concerns, social relationships, and self-acceptance (Cronbach's alphas 0.66–0.91). The sample size for the factor analysis was too small and, as a result, internal consistency was rated as indeterminate. Construct validity was assessed by formulating hypotheses *a priori* (except one) and were confirmed, which resulted in a positive rating. Floor or ceiling effects were assessed and 16% of the respondents scored the lowest possible

Table 2: Overview Included Internalized Stigma Instruments

Authors and Year	Country	Type of Instrument (name)	Language^	Target Population	Type of Stigma	No. of Subjects	No. of Items	Response Scale
HIV/AIDS Sowell *et al.*, 1997	USA	HIV Stigma Scale	English Arabic	PLWHA Rural areas	Perceived stigma Internalized stigma	82 Participants	13 Items	Four-point scale Not at all, rarely, sometimes and often
Fife & Wright, 2000	USA	Social Impact Scale (SIS)	English Russian	PLWHA Participants with cancer	Perceived stigma Enacted stigma Internalized stigma	206 Participants 130 PLWHA 76 Cancer participants	24 Items Social rejection (9) Financial insecurity (3) Internalized shame (5) Social isolation (6)	Four-point Likert scale
Berger, Ferrans & Lashley, 2001	USA	HIV Stigma Scale	English Portuguese Amharic Tamil Bahasa Malaysia	PLWHA	Perceived stigma Enacted stigma Internalized stigma	9 Participants for pretesting 318 Participants 139 Participants test – retest reliability	40 Items Personalized stigma (18) Disclosure concerns (8) Negative self-image (8) Concern with public attitudes (8)	Four-point Likert scale
Emlet, 2005	USA	HIV Stigma Scale (adapted version Sowell's HIV Stigma Scale)	English	PLWHA Older adults	Perceived stigma Internalized stigma	88 Participants	13 Items Distancing (4) Blaming (4) Discrimination (4) Undefined (1)	Four-point scale Not at all, rarely, sometimes and often
Bunn, Solomon, Miller & Forehand, 2007	UK	HIV Stigma Scale (adapted version Berger's HIV Stigma Scale)	English	PLWHA	Perceived stigma Enacted stigma Internalized stigma	157 Participants	32 Items Enacted stigma (11) Disclosure concerns (8) Negative self-image (7) Concern with public attitudes (6)	Four-point Likert scale
Holzemer *et al.*, 2007	Lesotho Malawi South Africa Swaziland Tanzania	HIV/AIDS Stigma Instrument-PLWA (HASI-P)	Sesotho Chichewa Afrikaans Setswana IsiZulu Seswati Swahili	PLWHA	Enacted stigma Internalized stigma	111 Participants focus group discussion 28 Participants interviews 217 Participants pilot testing 1477 Participants	33 Items Verbal abuse (8) Negative self-perception (5) Healthcare neglect (7) Social isolation (5) Fear of contagion (6) Workplace stigma (2)	Four-point scale Never, once or twice, several times, most of the time
Wright, Naar-King, Lam, Templin & Frey, 2007	USA	HIV Stigma Scale (adapted version Berger's HIV Stigma Scale)	English	PLWHA Youth	Perceived stigma Enacted stigma Internalized stigma	48 Participants	10 Items Personalized stigma (3) Disclosure (2) Negative self-image (3) Public attitudes (2)	Five-point Likert scale

(Continued on next page)

Table 2: *(Continued)*

Authors and Year	Country	Type of Instrument (name)	Language^	Target Population	Type of Stigma	No. of Subjects	No. of Items	Response Scale
Sayles *et al.*, 2008	USA	HIV internalized stigma measure	English	PLWHA	Internalized stigma	48 Participants focus group discussion 10 Participants cognitive interviews 202 Participants	28 Items Stereotypes (12) Disclosure concerns (5) Social relationships (6) Self acceptance (4)	Five-point scale None of the time, a little bit of the time, some of the time, most of the time, all of the time
Visser, Kershaw, Makin & Forsyth, 2008	South Africa	HIVInternalized stigma scale	English Sepedi Setswana Isizulu	PLWHA HIV-positive pregnant women	Internalized stigma	317 Participants	12 Items	Dichotomously Agree, disagree
Kalichman *et al.*, 2009	USA South Africa Swaziland	Internalized AIDS-related Stigma Scale (IA-RSS)	English Xhosa Afrikaans Swati	PLWHA	Internalized stigma	1068 Participants Cape Town 1090 Participants Swaziland 239 Participants Atlanta 181 Participants test–retest reliability	6 Items	Dichotomously Agree, disagree
Franke *et al.*, 2010	Peru	HIV Stigma Scale (adapted version Berger's HIV Stigma Scale)	Spanish	PLWHA Spanish speaking urban population Peru	Perceived stigma Enacted stigma Internalized stigma	130 Participants	21 Items Enacted stigma (5) Disclosure concerns (5) Negative self image (6) Concern with public attitudes (5)	Four-point scale Almost always, sometimes, rarely, almost never
Jimenez *et al.*, 2010	Puerto Rico	HIV Felt-Stigma Scale (HFSS) (adapted version Berger's HIV Stigma Scale)	Spanish	PLWHA Spanish Speaking Latino Population	Perceived stigma Enacted stigma Internalized stigma	40 Participants focus group 20 Participants grammar and editing process 156 Participants factor analysis 106 Participants construct validity 50 Participants test–retest analysis	17 Items Personalized stigma (5) Disclosure concerns (4) Negative self-image (5) Concern with public attitudes (3)	Four-point Likert scale
Tuberculosis HIV/AIDS								
Van Rie *et al.*, 2008	Thailand	HIV/AIDS and TB related stigma scale	English Thai Malayu	Persons affected by TB Persons affected by HIV	Internalized stigma Attitudes	Study 1 204 TB persons Study 2 480 TB persons	23 Items Community perspective toward TB (11) Patient perspective toward TB (12)	Four-point Likert scale

(Continued on next page)

Table 2: *(Continued)*

Authors and Year	Country	Type of Instrument (name)	Language^	Target Population	Type of Stigma	No. of Subjects	No. of Items	Response Scale
							21 Items Community perspective toward HIV/AIDS (11) Patient perspective toward HIV/AIDS (10)	
Leprosy								
Rensen *et al.*, 2010	India	Internalized Stigma of Mental Illness (ISMI) (Adapted version)	Tamil Bengali	Persons affected by leprosy	Internalized stigma Enacted stigma	806 Participants 165 controls	28 Items Alienation (6) Stereotype endorsement (6) Discrimination experience (5) Social withdrawal (6) Stigma resistance (5)	Four-point Likert scale
Mental illness								
Ritsher, Otilingam & Grajales, 2003	USA	Internalized Stigma of Mental Illness (ISMI)	English Croatian Yoruba Hebrew German Afrikaans Xhosa French Polish Spanish Swedish Turkish Greek Italian Czech Slovak	Persons affected by mental illness	Internalized stigma Enacted stigma	127 Participants	29 Items Alienation (6) Stereotype endorsement (7) Discrimination experience (5) Social withdrawal (6) Stigma resistance (5)	Four-point Likert scale
Corrigan, Watson & Barr, 2006	USA	Self-Stigma of Mental Illness Scale (SSMIS)	English	Persons affected by a mental illness	Internalized stigma	54 Participants study 1 60 Participants study 2	60 Items Stereotype awareness (15) Stereotype agreement (15) Self-concurrence (15) Self-esteem (15) (revised 40 items; 5 less on each subscale)	Nine-point Likert scale

(Continued on next page)

Table 2: *(Continued)*

Authors and Year	Country	Type of Instrument (name)	Language^	Target Population	Type of Stigma	No. of Subjects	No. of Items	Response Scale
Ersoy & Varan, 2007	Turkey	Internalized Stigma of Mental Illness (ISMI) (Validation of an Turkish version of the ISMI)	Turkish	Persons affected by a psychiatric disorder	Internalized stigma Enacted stigma	203 Participants	29 Items Alienation (6) Stereotype endorsement (7) Perceived discrimination (5) Stigma resistance (5) Social withdrawal (6)	Four-point Likert scale
Fung, Tsang, Corrigan, Lam & Cheng, 2007	China	Chinese Version of Self-stigma of Mental Illness Scale (CSSMIS) (Adapted version of the SSMIS)	Chinese	Persons affected by a mental illness	Internalized stigma	108 Participants 31 Participants test–retest reliability	60 Items Stereotype awareness (15) Stereotype agreement (15) Self-concurrence (15) Self-esteem (15)	Nine-point Likert scale
Pan, Chung, Fife & Hsiung, 2007	Taiwan	Social Impact Scale (SIS) (Further validation across health conditions and different country)	Chinese	Persons affected by mental illness HIV/AIDS and schizophrenia	Enacted stigma Perceived stigma Internalized stigma	580 Participants	24 Items Social rejection (9) Financial insecurity (3) Internalized shame (5) Social isolation (8)	Four-point Likert scale
Burgener & Berger, 2008	USA	Social Impact Scale (SIS) (Further validation in different health conditions)	English	Persons diagnosed with Alzheimer's dementia and Parkinson's disease	Enacted stigma Perceived stigma Internalized stigma	40 Participants	24 Items Social rejection (9) Financial insecurity (3) Internalized shame (5) Social isolation (8)	Four-point Likert scale
Kanter, Rusch & Brondino, 2008	USA	Depression Self-Stigma Scale (DSSS)	English	Persons affected by a mental illness Especially depression	Internalized stigma Perceived stigma	391 Participants	32 Items General self-stigma (9) Secrecy (9) Public stigma (4) Treatment stigma (4) Stigmatizing experiences (6)	Seven-point Likert scale
Moses, 2009	USA	Self-Stigma Scale	English	Persons affected by a mental illness Adolescents	Internalized stigma	60 Participants	12 Items 5 items Self-stigma 7 items Secrecy	Four-point Likert scale Four-point Likert scale Almost never, sometimes, frequently, very often

(Continued on next page)

Table 2: *(Continued)*

Authors and Year	Country	Type of Instrument (name)	Language^	Target Population	Type of Stigma	No. of Subjects	No. of Items	Response Scale
Barney, Griffiths, Christensen & Jorm, 2010	Australia	Self-Stigma of Depression Scale (SSDS)	English	Persons affected by a mental illness	Internalized stigma	37 Participants focus group discussion Study 1 408 Participants Study 2 330 Participants 1312 Participants	16 Items Shame (4) Self-blame (4) Help-seeking inhibition (4) Social inadequacy (4)	Five-point Likert scale
Mak & Cheung, 2010	China	Self-Stigma Scale (SSS)	English Chinese	Concealable minorities Persons affected by a mental illness, PLWHA etc.,	Internalized stigma	65 Participants focus group discussion 175 Participants mental illness 151 Participants mental illness	9 Items Self-stigmatizing cognitions (3) Self-stigmatizing affect (3) Self-denigration (3)	Four-point Likert scale
Cancer								
Esplen, Stuckless, Berk, Butler & Gallinger, 2009a	Canada	Familial Adenomatous Polyposis (FAP) Self-Concept Scale (FAPSC)	English	Persons diagnosed FAP	Self-concept Internalized stigma	132 Participants	23 Items Stigma (11) Self-esteem (7) Mastery (5)	Seven-point Likert scale
Esplen *et al.*, 2009b	Canada	BRCA Self-Concept Scale (BRCASC)	English	Women tested positively for BRCA 1/2 mutations	Self-concept Internalized stigma	35 Participants interviews/focus group discussion Study 1 115 Participants Study 2 126 Participants	23 Items Stigma (11) Self-esteem (7) Mastery (5)	Seven-point Likert scale
Petersen *et al.*, 2010	Denmark Canada Sweden	Self-Concept Scale for Lynch Syndrome	Danish Swedish English	Persons diagnosed with Lynch syndrome	Self-concept Internalized stigma	576 Participants	20 Items Stigma-vulnerability (15) Bowel symptom-related anxiety (5)	Seven-point Likert scale
Esplen *et al.*, 2011	Canada	Self-Concept Scale for Lynch Syndrome	English	Persons diagnosed with Lynch syndrome	Self-concept Internalized stigma	115 Participants	20 Items Stigma-vulnerability (15) Bowel symptom-related anxiety (5)	Seven-point Likert scale

(Continued on next page)

Table 2: *(Continued)*

Authors and Year	Country	Type of Instrument (name)	Language^	Target Population	Type of Stigma	No. of Subjects	No. of Items	Response Scale
Epilepsy Asthma								
Austin and Huberty, 1993	USA	Child Attitude Toward Illness Scale (CATIS)	English	Children affected by epilepsy or asthma	Internalized stigma	Study 1 50 Participants Study 2 269 Participants	13 Items	Five-point scale Various response options
Heimlich, Westbrook, Austin, Cramer & Devinsky, 2000	USA Canada	Child Attitude Toward Illness Scale (CATIS)	English	Adolescents affected by epilepsy	Internalized stigma	Study 1 189 Adolescents Study 2 178 Adolescent,	13 Items	Five-point scale Various response options
Obesity								
Durso & Latner, 2008	USA	Weight Bias Internalization Scale (WBIS)	English	Overweight or obese persons	Internalized stigma	198 Participants	11 Items	Seven-point Likert scale
Lillis, Luoma, Levin & Hayes, 2010	USA	Weight Self-Stigma Questionnaire (WSSQ)	English	Overweight or obese persons	Internalized stigma Perceived stigma	Study 1 85 Participants Study 2 84 Participants	12 Items Fear of enacted stigma (6) Weigh-related self-devaluation (6)	Five-point Likert scale
Generic								
Rao *et al.*, 2009	USA	Stigma Scale for Chronic Illness (SSCI)	English	Persons affected by a chronic illness	Internalized stigma Enacted stigma	511 Persons	24 Items Internalized stigma (13) Enacted stigma (11)	Five-point scale Never, rarely, sometimes, often, always

Notes: ^The overview provided concerning the available language versions of an instrument may be incomplete, due to the extensive body of literature.
PLWHA = persons living with HIV/AIDS.

Table 3: Overview of Quality Criteria Ratings Summarized Across All Studies Using the Instruments

Questionnaire*	Content Validity	Internal Consistency	Criterion Validity	Construct Validity	Reproducibility		Responsiveness	Floor or Ceiling Effects	Interpretability
					Agreement	Reliability			
HIV/AIDS									
HIV Stigma Scale (Sowell *et al.*, 1997)	+	?	NA	?	0	0	0	0	?
SIS (Fife & Wright 2000)	+	+	NA	?	0	0	0	0	0
HIV Stigma Scale (Berger *et al.*, 2001)	+	?	?	?	0	?	?	0	0
HASI-P (Holzemer *et al.*, 2007)	+	?	NA	?	0	0	0	0	0
Internalized HIV Stigma Measure (Sayles *et al.*, 2008)	+	?	NA	+	0	0	0	?	?
Stigma Scale (Visser *et al.*, 2008)	-	-	NA	?	0	0	0	0	0
IA-RSS (Kalichman *et al.*, 2009)	+	?	NA	?	0	?	0	0	?
Tuberculosis and HIV/AIDS									
TB-related Stigma Scale (Van Rie *et al.*, 2008)	+	?	NA	?	0	?	0	+	0
Mental Health									
ISMI (Ritsher *et al.*, 2003)	+	?	NA	+	0	?	0	+	?
(C)SSMIS (Corrigan *et al.*, 2006)	+	?	NA	?	0	?	0	0	0
SSS – S (Mak & Cheung, 2010)	+	?	?	?	0	0	0	0	?
Self-Stigma and Secrecy (Moses 2009)	?	?	NA	?	0	0	0	0	?
DSSS (Kanter *et al.*, 2008)	-	?	NA	?	0	0	0	0	?
SSDS (Barney *et al.*, 2010)	?	+	NA	?	0	-	0	0	?
Cancer									
BRCASC (Esplen *et al.*, 2009)	+	?	NA	?	0	0	0	+	?
FAPSC (Esplen *et al.*, 2009)	+	?	NA	?	0	0	0	?	?
Self-concept for Lynch Syndrome (Esplen *et al.*, 2011)	+	?	NA	?	0	0	0	0	0
Epilepsy and Asthma									
CATIS (Austin & Huberty, 1993)	+	+	NA	?	0	+	0	0	?
Obesity									
WBIS (Durso & Latner, 2008)	-	+	NA	?	0	0	0	0	0
WSSQ (Lillis *et al.*, 2010)	+	+	NA	?	0	?	?	0	0
Generic									
SSCI (Rao *et al.*, 2009)	+	-	NA	-	0	0	0	0	0

Notes: *Only the author(s) of the initial development study are included. The final rating is a summary of all ratings obtained across different studies.
Rating: + = positive rating; ? = indeterminate rating; – = negative rating; 0 = no information available NA, not applicable.
HASI-P = HIV/AIDS Stigma Instrument-PLWA; IA-RSS = Internalized Aids-Related Stigma Scale; ISMI = Internalized Stigma of Mental Illness; SIS = Social Impact Scale; (C)SSMIS = (Chinese) Self-Stigma of Mental Illness Scale; SSS-S = Self-Stigma Scale – Short; DSSS = Depression Self-Stigma Scale; SSCI = Stigma Scale for Chronic Illness; SSDS = Self-Stigma of Depression Scale; BRCASC = BRCA Self-Concept; FAPSC = FAP Self-Concept; CATIS = Child Attitude Towards Illness Scale; WBIS = Weight Bias Internalization Scale; WSSQ = Weight Self-Stigma Questionnaire.

score on the disclosure concerns subscale, resulting in an indeterminate rating (borderline; threshold >15%). On the other subscales, no floor or ceiling effects were identified (Sayles *et al.*, 2008).

Visser *et al.* (2008) developed a 12-item internalized stigma instrument. Content validity was rated negatively, because the actual target population was not involved during the item selection and generation process. Factor analysis (17 items, n=317) was applied and revealed two subscales: blame and judgement and interpersonal distancing, both with an alpha of 0.61. Construct validity was rated 'indeterminate', because the exact magnitude of the expected correlations was not specified (Visser *et al.*, 2008).

More recently, in 2009, Kalichman *et al.* developed the six-item Internalized AIDS-Related Stigma Scale (IA-RSS). Items were selected from existing instruments, after which draft items were tested with help of members of the target population. No factor analysis was applied, which resulted in an indeterminate rating, despite alphas of 0.73, 0.74 and 0.76, found across the countries (South-Africa, Swaziland and the USA). Hypotheses were formulated and were in the expected directions. Despite this, an indeterminate rating was given, because the absolute magnitudes of the *a priori* hypotheses were not defined. In addition, reliability was rated as indeterminate, because Pearson's correlation coefficients were used instead of weighted kappa or intraclass correlation coefficient s (Kalichman *et al.*, 2009). See Table 3.

Tuberculosis and HIV/AIDS

Few quantitative studies were identified that measured stigma and related constructs in tuberculosis (TB). Van Rie *et al.* (2008) simultaneously developed two scales to assess the stigma associated with tuberculosis and HIV/AIDS in Thailand. Stigma was assessed from two different perspectives, the community and the patient. Content validity was rated positively. A literature review was used as input for the initial item pool. Validity of these items was assessed with help of interviews and focus group discussions with members of the target population. Internal consistency was rated as indeterminate, because explanatory factor analysis was applied on an insufficient sample size (43 TB items, n=204; 41 HIV items, n=204). Cronbach's alphas for the different HIV/AIDS and tuberculosis subscales ranged from 0.82 to 0.92. Construct validity was rated indeterminate, as no exact magnitudes were defined for the hypothesized correlations. Reliability also was rated indeterminate, because a sample of only 15 respondents was used to assess test-retest reliability with Pearson's correlation. No floor or ceiling effects were identified, resulting in a positive rating (Van Rie *et al.*, 2008).

Mental Health

Ritsher *et al.* (2003) developed the 29-item Internalized Stigma of Mental Illness (ISMI) scale. An extensive literature review was conducted and in combination with focus group discussions and team meetings, this resulted in a pool of items (Ritsher *et al.*, 2003). No adequate factor analysis was applied to test the factor structure of the scale, resulting in an indeterminate rating. During the development study, all the alphas were found to be above the threshold of α=0.70, except for the stigma resistance subscale (α=0.58) (Ritsher *et al.*, 2003). Cronbach's alphas in the latest study ranged from 0.79 to 0.96 (Rensen *et al.*, 2010). Construct validity was rated positively. In the initial study no exact magnitude was provided in the *a priori* formulated hypotheses, whereas this was done adequately in the latest study, except for one hypothesis (Ritsher *et al.*, 2003; Rensen *et al.*, 2010). Test–retest reliability was assessed during the further validation study. However, due to miscommunication the time interval was 1–3 months instead of 1–2 weeks, resulting in an indeterminate rating. No floor or ceiling effects were identified (Rensen *et al.*, 2010). A Turkish version of the Internalized Stigma of Mental Illness Scale was validated in 2007 (Ersoy & Varan, 2007). Content validity was rated indeterminate, because members of the target population were not involved during the translation procedure. Internal consistency and construct validity were both rated as indeterminate. No factor analysis was applied and hypotheses were not formulated *a priori*. Cronbach's alphas ranged from 0.63 to 0.87.

The Self Stigma of Mental Illness Scale (SSMIS) was developed to assess the levels of internalized and perceived stigma of persons diagnosed with a mental illness (Corrigan *et al.*, 2006). Items were generated, after which they were tested during a focus group discussion with respondents resulting in a positive rating. Finally, 60 items were retained in four different subscales each with 15 items, but without the application of factor analysis. Cronbach's alphas ranged from 0.72 to 0.91. Construct validity was assessed; however, no specific *a prior* hypotheses were formulated, resulting in an indeterminate rating. In addition, reliability was also rated as indeterminate. Fifty-four respondents were assessed for a second time within 1 week. The majority of the outcomes were above the threshold of α=0.70 (range 0.68–0.82), but we were unable to identify the exact statistical method they applied (Corrigan *et al.*, 2006). A Chinese version of the Stigma of Mental Illness scale was validated by Fung *et al.* (2007) and during this study factor analysis was applied (60 items, n=108). This suggested a five-factor scale (Cronbach's alpha's 0.82–0.90). Also in this study, no specific hypotheses with exact magnitudes were formulated to assess construct validity.

Test–retest reliability showed good results, with an ICC range of 0.71–0.81, but the sample consisted of only 31 respondents, resulting in an indeterminate rating (Fung *et al.*, 2007).

The short version nine-item Self-Stigma Scale (SSS-S) was developed for use in minorities with concealable conditions such as people with a mental illness, people living with HIV/AIDS, and immigrants (Mak & Cheung, 2010). In their study, samples of immigrants and people diagnosed with a mental illness were included. We will focus only on the results found for the latter, because this review concerns health-related stigma. Items were selected in close collaboration with members of the target population. Internal consistency was rated as indeterminate, because sample one (mental health consumers) comprised only 175 participants when explanatory factor analysis was applied on the 39-item version of the scale. Cronbach's alphas ranged from 0.81 to 0.84. Criterion validity was rated as indeterminate because no adequate comparison was made between the long and abbreviated version of the scale. Construct validity was assessed by investigating relationships between internalized stigma and other closely related constructs (e.g., self-esteem, self-efficacy), which were all negatively associated. However, no adequate hypotheses were formulated, resulting in an indeterminate rating (Mak & Cheung, 2010).

Moses (2009) validated several stigma measures adapted for use in adolescents diagnosed with a mental illness. Due to the focus of our study, we will describe only the results found for the further validation of the five-item self-stigma (Austin, MacLeod, Dunn, Shen & Perkins, 2004) and the seven-item secrecy scale (Link, Mirotznik & Cullen, 1991; Link, Struening, Rahav, Phelan & Nuttbrock, 1997; Fife & Wright, 2000). The majority of the items were derived from existing measures without involvement of members of the target population. This resulted in an indeterminate rating. The revised Self-Stigma Scale showed an alpha of 0.81 and the secrecy scale had an alpha of 0.84. Despite this, internal consistency was rated as indeterminate, because the sample size for factor analysis was not sufficient. According to the authors, construct validity was demonstrated by significant positive correlations among the stigma subscales ($r = 0.29–0.64$) as well as with other measures of self-concept and depression. However, no specific hypotheses were formulated, resulting in an indeterminate rating (Moses, 2009).

The Depression Self-Stigma Scale (DSSS) was developed to identify and measure distinct constructs associated with depression self-stigma (Kanter *et al.*, 2008). Content validity of this scale was rated negatively because members of the target population were not involved during item generation or selection. The 32-item scale encompasses five subscales: general self-stigma, secrecy, public stigma, treatment stigma, and stigmatizing experiences. Cronbach's alphas were good (ranging from 0.79 to 0.93), but factor analysis was not applied on an adequate sample size (59 items, n=391). Construct validity was rated also as indeterminate, because the exact magnitudes of the hypotheses formulated were not specified *a priori*. Means and standard deviations of the stigma scores were presented for gender and ethnicity. However, no minimally important change was defined, so also interpretability was rated as indeterminate (Kanter *et al.*, 2008).

Recently, Barney *et al.* (2010) developed the Self-Stigma of Depression Scale (SSDS) with input from focus group discussions including persons with and without a history of depression, and a literature review conducted by the researchers. However, content validity was rated as indeterminate because it is unclear how the final 16-item scale was established. Internal consistency was rated positive. Factor analysis was applied on an adequate sample size (sample 1; 19 items, n=408; sample 2; 25 items, n=330) and Cronbach's alphas ranged from 0.78 to 0.83 for the four sub-scales. Construct validity was rated as indeterminate, because no specific hypotheses were formulated *a priori*. Test–-retest reliability was assessed in 151 respondents after two months. Intraclass correlation coefficients varied from 0.49 to 0.63 and this resulted in a negative rating. Interpretability was rated as indeterminate. Stigma scores were stratified for gender and depression experiences, but no minimally important change was defined (Barney *et al.*, 2010) (See Table 3).

Cancer

A few studies have measured stigma related to cancer. In 2009, two scales were developed to measure self-concept in persons diagnosed with familial adenomatous polyposis and women who tested positive for BRCA1/2 mutation (Esplen *et al.*, 2009a,b). Both scales showed good content validity, with members of the target population actively involved during the item generation and selection process, and a clear construct was described that was to be measured with the scale (Esplen *et al.*, 2009a,b). Factor analysis was completed on an adequate sample in the development study of the BRCA Self-Concept Scale study (25 items, n=241), however, an alpha of 0.68 was found for one of the three subscale (threshold α=0.70), resulting in an indeterminate rating (Esplen *et al.*, 2009b). Factor analysis was applied on the Familial Adenomatous Polyposis Self-Concept Scale on an insufficient sample size (23 items,

n=132), resulting in an indeterminate rating. Cronbach's alphas ranged from 0.73 to 0.89 (Esplen *et al.*, 2009a). Construct validity was rated as indeterminate for both scales because the exact magnitudes of the hypotheses were not defined. No floor or ceiling effects were examined for the BRCA Self-Concept Scale. For the Familial Adenomatous Polyposis Self-Concept Scale this rating was indeterminate, because no conclusion could be drawn based on the results presented (Esplen *et al.*, 2009a,b).

Recently, Esplen *et al.* developed the 20-item self-concept scale for Lynch Syndrome (Esplen *et al.*, 2011). During the initial study, individual interviews and focus group discussions with members of the target population and genetic counsellors were conducted which formed the basis for identification of items for uptake in the scale. This resulted in a positive rating for content validity. Hypotheses to assess construct validity were formulated *a priori*; however the exact magnitude of the correlations was not defined in the majority of these hypotheses, resulting in an indeterminate rating. Cronbach's alphas for the subscales were 0.83 and 0.92, respectively for the bowel symptom related anxiety subscale and the stigma/vulnerability subscale. A further study was designed to validate the scale for administration in Denmark, Sweden, and Canada (Petersen *et al.*, 2011). Factor analysis was applied (20 items, n=576) and revealed three factors: stigma and vulnerability, bowel symptom-related anxiety, and positive future-directed statements. During this study, no alphas for the subscales were reported (total scale α=0.93) so internal consistency was rated as indeterminate (Petersen *et al.*, 2011).

Epilepsy and Asthma

The Child Attitude Towards Illness Scale (CATIS) was initially developed to assess attitudes of children aged 8–12 years towards their illness and further validated for use in adolescents (Austin & Huberty, 1993; Heimlich *et al.*, 2000). Children with asthma (n=133) and epilepsy (n=136) were included in the former study; in the latter the adolescent sample consisted of 197 epilepsy patients (Austin & Huberty, 1993; Heimlich *et al.*, 2000). The combination of the initial development study and the further validation of the scale resulted in three positive ratings on the quality criteria content validity, internal consistency and reliability. In both studies the Cronbach's alphas were above the threshold of α=0.70, respectively 0.77 and 0.80. Factor analysis was applied during the initial study on an adequate sample size (13 items, n=269). Reliability investigated in the study by Heimlich *et al.* (2000) was good, with an Intraclass correlation coefficient of 0.77 (n=166). The other criteria were rated as no information available or indeterminate.

Obesity

Durso & Latner (2008) created the 11- item Weight Bias Internalization Scale (WBIS) for use in populations of overweight and obese persons. This scale received a negative rating for content validity because members of the target population were not involved in the scale development. A positive rating was given for internal consistency because factor analysis was applied on a sufficient sample size (13 items, n=198) and the Cronbach's alpha was 0.90. Construct validity was rated as indeterminate because no specific hypotheses were formulated in advance (Durso & Latner, 2008).

The 12-item Weight Self-Stigma Questionnaire has two subscales measuring fear of enacted stigma and weight-related self-devaluation. They were derived after explanatory factor analysis was performed on an adequate sample size (22 items, n=169) (Lillis *et al.*, 2010). Cronbach's alphas were 0.87 and 0.81, respectively. Content validity was rated 'good' due to members of the target population being involved during the item selection process and a focus group discussion between researchers during the item development process. Construct validity was rated as indeterminate. Several instruments were included to assess associations with the stigma scores. However, no specific hypotheses were formulated *a priori*. Forty-four respondents filled in the scale for a second time after a 3-month time interval and this resulted in an intraclass correlation coefficient of 0.79 for the total scale and 0.80 and 0.62 for the subscales. Reliability was rated indeterminate, due to the small sample size (Lillis *et al.*, 2010). Further details are given in Table 3.

Generic

A recently developed instrument, the Stigma Scale for Chronic Illness (SSCI), assesses internalized and experienced stigma across chronic illnesses (Rao *et al.*, 2009). This 24-item measure has been tested in a study sample of 511 respondents suffering from several health conditions including multiple sclerosis, epilepsy and Alzheimer's disease. Content validity was rated as positive. The concepts were clearly stated, as well as involvement of the target population and experts during the scale development process. However, internal consistency was rated as negative. The Cronbach's alpha (0.97) indicated that the scale might be shortened. Construct validity was also rated

negatively since not all hypothesized correlations, such as the relationship of stigma with pain, psychologically distress and performance status, were as expected (Rao *et al.*, 2009).

Discussion

We reviewed 21 instruments developed to measure internalized stigma and several related constructs (such as anticipated or perceived stigma or experienced stigma) from 33 studies that were identified following a systematic review. We used Livingston & Byod's (2010) definition of internalized stigma, as a 'subjective process, embedded within a socio-cultural context, which may be characterized by negative feelings (about self), maladaptive behaviour, identity transformation or stereotype endorsement resulting from an individual's experiences, perceptions, or anticipation of negative social reaction on the basis of their health condition's shown in Figure 1, but would like to provide a critical note. According to their definition, stereotype endorsement can be placed on the manifestation level; however, we suggest that internalized stigma encompasses stereotype endorsement so inserted bidirectional arrows between the process and manifestation level and between the constructs on the levels. For instance, people who have experienced a lot of stigmatization, also tend to anticipate it more frequently, and vice versa. We suggest a possible further modification – the insertion of a psychological mechanism level between levels 2 and 3. Before a person experiences internalized stigma, the person must be aware, agree and apply the manifestations of public stigma to him or herself (Corrigan et al., 2006; Watson, Corrigan, Larson & Sells, 2007). These relationships would need to be empirically confirmed in future research.

Only two of the 21 instruments reviewed, the Child Attitude Towards Illness Scale and the Internalized Stigma of Mental Illness scale, received three positive ratings (Austin & Huberty, 1993; Heimlich *et al.*, 2000; Ritsher *et al.*, 2003; Fung *et al.*, 2007; Rensen *et al.*, 2010). Six instruments scored two positive ratings; the Tuberculosis Stigma Scale (Van Rie *et al.*, 2008), the Weight Self-stigma Questionnaire (Lillis *et al.*, 2010), the Breast Cancer Self-concept Scale (Esplen *et al.*, 2009b), the Stigma of Depression Scale (Barney *et al.*, 2010), the Internalized HIV Stigma Measure (Sayles *et al.*, 2008) and the Social Impact Scale (Fife & Wright, 2000). The majority of these positive ratings concerned content validity, internal consistency, reliability, and floor and ceiling effects.

The Stigma Scale received two negative ratings, one for internal consistency, because the Cronbach's alpha was only 0.61 and one for content validity, because members of the target population were not involved during its development (Visser *et al.*, 2008). The Stigma Scale for Chronic Illness also received two negative ratings (Rao *et al.*, 2009). Three instruments received one negative rating: the Weight Bias Internalization Scale (Durso & Latner, 2008), the Self-stigma of Depression Scale (Barney *et al.*, 2010) and the Depression Self-stigma Scale (Kanter *et al.*, 2008). The majority of the negative ratings given in the present review were for content validity and internal consistency. The negative ratings for content validity were most often due to the target population not being involved during item generation and selection as this was one of the requisites for a positive rating for content validity. Internal consistency was often rated as negative due to insufficient Cronbach's alphas (≤ 0.70) reported for the subscales.

Except the Internalized Stigma of Mental Illness Scale, all instruments received at least three or more zero ratings, indicating missing information. This does not necessarily indicate that these instruments are of poor quality, but that additional assessment of these psychometric properties is necessary. Of the 21 instruments included, 12 received at least three indeterminate ratings. The majority of the indeterminate ratings were given for construct validity (18 instruments) and internal consistency (13 instruments). The large number of zero and indeterminate ratings may be due to the fact that the criteria of Terwee *et al.* (2007) are fairly new and not incorporated thoroughly in the studies assessed.

Many studies tested construct validity using correlations with related constructs. According to Terwee *et al.* (2007) *specific* hypotheses should be formulated *a priori* and at least 75% of the results should be in accordance with the expectations. The term *specific* refers to the adequate description of a hypothesis; namely the inclusion of the direction as well as the expected magnitude of the correlation (Terwee *et al.*, 2007; Mokkink *et al.*, 2010b). In most cases, the magnitude of the expected association was not specified resulting in an indeterminate rating. If a hypothesis states that a 'moderate or small correlation' is expected, this is open to interpretation so is also inadequate. An expected direction and a range should be specified. For example: 'A positive correlation of 0.40–0.60, as measured with Pearson's correlation coefficient'.

Internal consistency was also often rated as 'indeterminate'. Terwee *et al.* (2007) stated that a positive rating for internal consistency should only be given if the sample size for factor analysis is at least 7 times the number of items and is 100 or more. The initial item pool often exceeds 50 items. As a result, the study sample should

consist of at least 350 respondents. In the majority of the studies included, sample sizes were smaller, resulting in an indeterminate rating. In addition, Cronbach's alphas should exceed 0.70. Especially the first requisite is an important barrier.

Criterion validity is also an important criterion. According to the Delphi panel that was used in the Consensus-based Standards for the selection of health Measurement Instruments study, no gold standard exists in the area of health-related patient-reported outcome instruments that can be used for the assessment of criterion validity (Mokkink *et al.*, 2010a). One exception is the comparison between an abbreviated version of an instrument and the longer original version. In this case, the longer version of the scale functions as the gold standard (Mokkink *et al.*, 2010a). In our review, we were only able to assess criterion validity thoroughly for the widely applied HIV Stigma Scale developed by Berger and colleagues (2001). For the most other instruments our rating was 'not applicable'.

A recent review published by Magasi & Post (2010) used the same quality criteria framework to assess instruments measuring (social) participation. They included 8 instruments that were primarily based on the International Classification of Functioning, Disability and Health (World Health Organization, 2001; Magasi & Post, 2010) and recommended further validation. The psychometric properties that were addressed showed moderate to good results, however, often information necessary for the assessment of certain psychometric properties was lacking (World Health Organization, 2001; Magasi & Post, 2010). Brohan, Slade, Clement & Thornicroft (2010) used an adapted version of the framework of Terwee *et al.* (2007). They assessed mental health stigma instruments only on content validity, internal consistency, construct validity, test–retest reliability and floor or ceiling effects (Brohan *et al.*, 2010). None of the instruments scored positively on all the ratings and several instruments were identified for further validity and reliability testing (Brohan *et al.*, 2010). We also found that psychometric testing was often not done adequately and needs further attention. In addition, Brohan and colleagues excluded rating agreement, responsiveness, and interpretability, because in most studies no information was presented to assess these. This corresponds with the results found in our review. Agreement and responsiveness received the least attention, whereas content validity, construct validity and internal consistency were most often studied. Magasi & Post (2010) also excluded the assessment of agreement and interpretability and also concluded that almost no information was presented to assess the criterion 'responsiveness'.

Two instruments included in the review by Brohan and colleagues, the Depression Self-Stigma Scale and the Self-stigma of Mental Illness Scale, were rated differently in our review. Brohan *et al.* gave the former an indeterminate rating for content validity and internal consistency, and a positive rating for construct validity. We gave a negative rating for content validity and indeterminate ratings for internal consistency and construct validity. The discrepancy for the latter occurred due to the *specific* description of construct validity that we applied. Based on the extended description concerning the item generation and selection provided in the article, we concluded that members of the target population were not involved and therefore rated content validity negatively. We rated internal consistency as 'indeterminate', because the sample size for the factor analysis was insufficient (e.g., 59 items, n=391). Brohan *et al.* assigned the Self-stigma of Mental Illness Scale a positive rating for construct validity and test-retest validity. We rated construct validity as 'indeterminate' due to the extended description mentioned earlier. Also test–retest reliability was rated as 'indeterminate' because we could not identify the statistical method applied for testing this criterion.

Based on previous research and our findings, we consider there is a need for a short testing protocol to guide researchers on how to design an appropriate instrument development studies, how to address particular psychometric properties, and on the preferred statistical methods for testing these. Such a protocol would contribute to the development of high quality instruments that are necessary to perform good research and set up adequate interventions and evaluate their impact. It is important to stress the need to assess the validity of newly developed instruments and before an instrument is applied in a different culture, health condition, or study population.

In conclusion, using current standard quality criteria for psychometric testing, we have provided an overview of the best-validated instruments to measure internalized stigma currently available. Twenty-one instruments were assessed and for most, essential information needed to rate the majority of the psychometric properties was lacking. The properties that could be assessed showed moderate to good results. Of all the instruments included, the Child Attitude Towards Illness Scale and the Internalized Stigma of Mental Illness Scale showed the best results, with three positive ratings. Further validation of all the instruments is recommended.

Acknowledgment

We would like to thank Prof. Dr. Matthias Angermeyer for his comments on earlier drafts of this paper.

References

Adewuya, A. O., Owoeye, A. O., Erinfolami, A. O., & Ola, B. A. (2010). Correlates of self-stigma among outpatients with mental illness in Lagos, Nigeria. *International Journal of Social Psychiatry*, http://dx.doi.org/10.1177/0020764010363522.

Arole, S., Premkumar, R., Arole, R., Maury, M., & Saunderson, P. (2002). Social stigma: A comparative qualitative study of integrated and vertical care approaches to leprosy. *Leprosy Review, 73,* 186–196. http://dx.doi.org/10.1093/jpepsy/18.4.467.

Austin, J. K., & Huberty, T. J. (1993). Development of the Child Attitude Toward Illness Scale. *Journal of Pediatric Psychology, 18,* 467–480. http://dx.doi.org/10.1093/jpepsy/18.4.467.

Austin, J. K., MacLeod, J., Dunn, D. W., Shen, J., & Perkins, S. M. (2004). Measuring stigma in children with epilepsy and their parents: Instrument development and testing. *Epilepsy Behavior, 5,* 472–482. http://dx.doi.org/10.1016/j.yebeh.2004.04.008.

Barney, L. J., Griffiths, K. M., Christensen, H., & Jorm, A. F. (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. http://dx.doi.org/10.1002/mpr.325.

Berger, B. E., Ferrans, C. E., & Lashley, F. R. (2001). Measuring stigma in people with HIV: Psychometric assessment of the HIV stigma scale. *Research in Nursing & Health, 24,* 518–529. http://dx.doi.org/10.1002/nur.10011.

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.* http://dx.doi.org/10.1186/1472-6963-10-80.

Bunn, J. Y., Solomon, S. E., Miller, C., & Forehand, R. (2007). Measurement of stigma in people with HIV: A reexamination of the HIV stigma scale. *Aids Education and Prevention, 19,* 198–208. http://dx.doi.org/10.1521/aeap.2007.19.3.198.

Burgener, S. C., & Berger, B. (2008). Measuring perceived stigma in persons with progressive neurological disease: Alzheimer's dementia and Parkinson's disease. *Dementia, 7,* 31–53. http://dx.doi.org/10.1177/1471301207085366.

Corrigan, P. W. (1998). The impact of stigma on severe mental illness. *Cognitive and Behavioral Practice, 5,* 201–222. http://dx.doi.org/10.1016/S1077-7229(98)80006-0.

Corrigan, P. W., Watson, A. C., & Barr, L. (2006). The self-stigma of mental illness: Implications for self-esteem and self-efficacy. *Journal of Social and Clinical Psychology, 25,* 875–884. http://dx.doi.org/10.1521/jscp.2006.25.8.875.

Dinos, S., Stevens, S., Serfaty, M., Weich, S., & King, M. (2004). Stigma: The feelings and experiences of 46 people with mental illness. Qualitative study. *British Journal of Psychiatry, 184,* 176–181. http://dx.doi.org/10.1192/bjp.184.2.176.

Durso, L. E., & Latner, J. D. (2008). Understanding self-directed stigma: Development of the weight bias internalization scale. *Obesity (Silver, Spring), 16 Suppl 2,* S80-S86. http://dx.doi.org/10.1038/oby.2008.448.

Emlet, C. (2005). Measuring stigma in older and younger adults with HIV/AIDS: An analysis of an HIV stigma scale and initial exploration of subscales. *Research on Social Work Practice, 15,* 291–300. http://dx.doi.org/10.1177/1049731504273250.

Ersoy, M. A., & Varan, A. (2007). Reliability and validity of the Turkish version of the Internalized Stigma of Mental Illness Scale. *Turk Psikiyatri Dergisi, 18,* 163–171.

Esplen, M. J., Stuckless, N., Berk, T., Butler, K., & Gallinger, S. (2009a). The FAP self-concept scale (adult form). *Family Cancer, 8,* 39–50. http://dx.doi.org/10.1007/s10689-008-9204-x.

Esplen, M. J., Stuckless, N., Hunter, J., Liede, A., Metcalfe, K., Glendon, G., Narod, S., Butler, K., Scott, J., Irwin, E. (2009b). The BRCA Self-Concept Scale: A new instrument to measure self-concept in BRCA1/2 mutation carriers. *Psychooncology, 18,* 1216–1229. http://dx.doi.org/10.1002/pon.1498.

Esplen, M. J., Stuckless, N., Gallinger, S., Aronson, M., Rothenmund, H., Semotiuk, K., Stokes, J., Way, C., Green, J., Butler, K., Petersen, H.V., Wong, J. (2011). Development and validation of an instrument to measure the impact of genetic testing on self-concept in Lynch syndrome. *Clinical Genetics, 80,* 415–423. http://dx.doi.org/10.1111/j.1399-0004.2011.01770.x.

Fife, B. L., & Wright, E. R. (2000). The dimensionality of stigma: A comparison of its impact on the self of persons with HIV/AIDS and cancer. *Journal of Health and Social Behavior, 41,* 50–67. http://dx.doi.org/10.2307/2676360.

Franke, M. F., Munoz, M., Finnegan, K., Zeladita, J., Sebastian, J. L., Bayona, J. N., Shin, S.S. (2010). Validation and abbreviation of an HIV stigma scale in an adult Spanish-Speaking Population in Urban Peru. *Aids and Behavior, 14,* 189–199. http://dx.doi.org/10.1007/s10461-008-9474-1.

Fung, K. M. T., Tsang, H. W. H., Corrigan, P. W., Lam, C. S., & Cheng, W. M. (2007). Measuring self-stigma of mental illness in China and its implications for recovery. *International Journal of Social Psychiatry, 53,* 408–418. http://dx.doi.org/10.1177/0020764007078342.

Heimlich, T. E., Westbrook, L. E., Austin, J. K., Cramer, J. A., & Devinsky, O. (2000). Brief report: Adolescents' attitudes toward epilepsy: Further validation of the Child Attitude Toward Illness Scale (CATIS). *Journal of Pediatric Psychology, 25,* 339–345. http://dx.doi.org/10.1093/jpepsy/25.5.339.

Holzemer, W. L., Uys, L. R., Chirwa, M. L., Greeff, M., Makoae, L. N., Kohi, T. W., Dlamini, P.S., Stewart, A.J., Mullan, J., Phetlhu, R.D., Wantland, D., Durrheim, K. (2007). Validation of the HIV/AIDS stigma instrument-PLWA (HASI-p). *Aids Care-Psychological and Socio-Medical Aspects of Aids/Hiv, 19,* 1002–1012. http://dx.doi.org/10.1080/09540120701245999.

Jimenez, J. C., Puig, M., Ramos, J. C., Morales, M., Asencio, G., Sala, A. C., Castro, E., Velez Santori, E., Santiago, L., Zorilla, C. (2010). Measuring HIV felt stigma: A culturally adapted scale targeting PLWHA in Puerto Rico. *Aids Care-Psychological and Socio-Medical Aspects of Aids/Hiv, 22,* 1314–1322. http://dx.doi.org/10.1080/09540121003758481.

Kalichman, S. C., Simbayi, L. C., Cloete, A., Mthembu, P. P., Mkhonta, R. N., & Ginindza, T. (2009). Measuring AIDS stigmas in people living with HIV/AIDS: The Internalized AIDS-Related Stigma Scale. *AIDS Care, 21,* 87–93. http://dx.doi.org/10.1080/09540120802032627.

Kanter, J. W., Rusch, L. C., & Brondino, M. J. (2008). Depression self-stigma: A new measure and preliminary findings. *Journal of Nervous and Mental Disease, 196,* 663–670. http://dx.doi.org/10.1097/NMD.0b013e318183f8af.

Lillis, J., Luoma, J. B., Levin, M. E., & Hayes, S. C. (2010). Measuring weight self-stigma: The weight self-stigma questionnaire. *Obesity (Silver Spring), 18,* 971–976. http://dx.doi.org/10.1038/oby.2009.353.

Link, B. G., Mirotznik, J., & Cullen, F. T. (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. http://dx.doi.org/10.2307/2136810.

Link, B. G., Struening, E. L., Rahav, M., Phelan, J. C., & Nuttbrock, L. (1997). On stigma and its consequences: Evidence from a longitudinal study of men with dual diagnoses of mental illness and substance abuse. *Journal of Health and Social Behavior, 38,* 177–190. http://dx.doi.org/10.1146/annurev.soc.27.1.363.

Livingston, J. D., & Boyd, J. E. (2010). Correlates and consequences of internalized stigma for people living with mental illness: A systematic review and meta-analysis. *Social Science & Medicine, 71,* 2150–2161. http://dx.doi.org/10.1016/j.socscimed.2010.09.030.

Magasi, S., & Post, M. W. (2010). A comparative review of contemporary participation measures' psychometric properties and content coverage. *Archives of Physical Medicine and Rehabilitation, 91,* S17–S28. http://dx.doi.org/10.1016/j.apmr.2010.07.011.

Mak, W. W., & Cheung, R. Y. (2010). Self-stigma among concealable minorities in Hong Kong: Conceptualization and unified measurement. *American Journal of Orthopsychiatry, 80,* 267–281. http://dx.doi.org/10.1111/j.1939-0025.2010.01030.x.

Mokkink, L. B., Terwee, C. B., Knol, D. L., Stratford, P. W., Alonso, J., Patrick, D. L., Bouter, L.M., de Vet, H.C.W. (2010a). The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. *BMC Medical Research Methodology, 10,* 22. http://dx.doi.org/10.1186/1471-2288-6-2.

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L.M., de Vet, H.C.W. (2010b). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology, 63,* 737–745. http://dx.doi.org/10.1016/j.jclinepi.2010.02.006.

Moses, T. (2009). Stigma and self-concept among adolescents receiving mental health treatment. *Am.J Orthopsychiatry, 79,* 261–274. http://dx.doi.org/10.1037/a0015696.

Pan, A. W., Chung, L., Fife, B. L., & Hsiung, P. C. (2007). Evaluation of the psychometrics of the Social Impact Scale: A measure of stigmatization. *International Journal of Rehabilitation Research, 30,* 235–238. http://dx.doi.org/10.1097/MRR.0b013e32829fb3db.

Petersen, H. V., Domanska, K., Bendahl, P. O., Wong, J., Carlsson, C., Bernstein, I., Esplen, M.J., Nilbert, M. (2011). Validation of a Self-Concept Scale for Lynch Syndrome in Different Nationalities. *Journal of Genetic Counselling.* http://dx.doi.org/10.1007/s10897-011-9349-x.

Rao, D., Choi, S. W., Victorson, D., Bode, R., Peterman, A., Heinemann, A., Cella, D. (2009). Measuring stigma across neurological conditions: The development of the stigma scale for chronic illness (SSCI). *Quality of Life Research, 18,* 585–595. http://dx.doi.org/10.1007/s11136-009-9475-1.

Rensen, C., Bandyopadhyay, S., Gopal, P. K., & Van Brakel, W. H. (2010). Measuring leprosy-related stigma – a pilot study to validate a toolkit of instruments. *Disability and Rehabilitation, 33,* 711–719. http://dx.doi.org/10.3109/09638288.2010.506942.

Ritsher, J. B., Otilingam, P. G., & Grajales, M. (2003). Internalized stigma of mental illness: Psychometric properties of a new measure. *Psychiatry Research, 121,* 31–49. http://dx.doi.org/10.1016/j.psychres.2003.08.008.

Sayles, J. N., Hays, R. D., Sarkisian, C. A., Mahajan, A. P., Spritzer, K. L., & Cunningham, W. E. (2008). Development and psychometric assessment of a multidimensional measure of internalized HIV stigma in a sample of HIV-positive adults. *Aids and Behavior, 12,* 748–758. http://dx.doi.org/10.1007/s10461-008-9375-3.

Sorsdahl, K. R., Mall, S., Stein, D. J., & Joska, J. A. (2011). The prevalence and predictors of stigma amongst people living with HIV/AIDS in the Western Province. *AIDS Care, 23,* 680–685. http://dx.doi.org/10.1080/09540121.2010.525621.

Sowell, R. L., Lowenstein, A., Moneyham, L., Demi, A., Mizuno, Y., & Seals, B. F. (1997). Resources, stigma, and patterns of disclosure in rural women with HIV infection. *Public Health Nursing, 14,* 302–312. http://dx.doi.org/10.1111/j.1525–1446.1997.tb00379.x.

Stevelink, S. A. M., Van Brakel, W. H., & Augustine, V. (2011). Stigma and social participation in Southern India: Differences and commonalities among persons affected by leprosy and persons living with HIV/AIDS. *Psychology, Health and Medicine, 16,* 695–707. http://dx.doi.org/10.1080/13548506.2011.555945.

Terwee, C. B., Bot, S. D. M., De Boer, M. R., van der Windt, D. A. W. M., Knol, D. L., Dekker, J., Bouter, L.M., de Vet, H.C.W. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology, 60,* 34–42. http://dx.doi.org/10.1016/j.jclinepi.2006.03.012.

Van Brakel, W. H. (2006). Measuring health-related stigma – a literature review. *Psychology, Health and Medicine, 11,* 307–334. http://dx.doi.org/10.1080/13548500600595160.

Van Rie, A., Sengupta, S., Pungrassami, P., Balthip, Q., Choonuan, S., Kasetjaroen, Y., Strauss, R.P., Chongsuvivatwong, V. (2008). Measuring stigma associated with tuberculosis and HIV/AIDS in southern Thailand: Exploratory and confirmatory factor analyses of two new scales. *Tropical Medicine & International Health, 13,* 21–30. http://dx.doi.org/10.1111/j.1365-3156.2007.01971.x.

Visser, M. J., Kershaw, T., Makin, J. D., & Forsyth, B. W. C. (2008). Development of parallel scales to measure HIV-Related stigma. *Aids and Behavior, 12,* 759–771. http://dx.doi.org/10.1007/s10461-008-9363-7.

Watson, A. C., Corrigan, P., Larson, J. E., & Sells, M. (2007). Self-stigma in people with mental illness. *Schizophrenia Bulletin, 33,* 1312–1318. http://dx.doi.org/10.1093/schbul/sbl076.

World Health Organization (2001). *International Classification of Functioning, Disability and Health (ICF).* Geneva: WHO.

Wright, K., Naar-King, S., Lam, P., Templin, T., & Frey, M. (2007). Stigma scale revised: Reliability and validity of a brief measure of stigma for HIV plus youth. *Journal of Adolescent Health, 40,* 96–98.