Reliability and validity of the Screening Tool for Assessment of Psychosocial Problems

Abstract

Background: While working in a multidisciplinary team, many a time there is lack of clarity or connivance who will work on what issue and how to decide about a referral to another team member, particularly to the psychiatric social worker. So, a need was felt to develop a screening tool for the same purpose. **Aim:** To develop a screening tool for the assessment of psychosocial problems, and to test its reliability and validity. **Methods:** A 12-item scale was developed following the scientific tool development steps. Content validity of the tool was established by accepting more than 70% validity index for each item. To test the reliability, it was applied to 100 family members of persons with mental illness. **Results:** Statistical analyses showed good internal consistency by Cronbach's alpha (r=0.76) and Spearman-Brown prophecy formula (r=0.73). Scores of all the 12 individual items significantly correlated with the total score of the tool indicating acceptable sensitivity. A significant correlation was found among the scores of items indicating good construct validity of the tool. **Conclusion:** The Screening Tool for Assessment of Psychosocial Problems (STAPP) is a brief and simple to use instrument, which has acceptable psychometric properties (valid and reliable). It is suitable for screening of the psychosocial problems of family members of person with a mental illness. It can be used for screening in routine clinical practice.

Keywords: Referral. Psychiatric Social Work. Assessment.

Kamlesh Kumar Sahu, Bir Singh Chavan, Chandra Bala, Shikha Tyagi

Department of Psychiatry, Government Medical College & Hospital, Chandigarh, India

Correspondence: Dr. Kamlesh Kumar Sahu, Level-5 Block-D, Department of Psychiatry, Government Medical College & Hospital, Sector 32, Chandigarh-160030, India. withkamlesh@gmail.com

INTRODUCTION

It has been recognised that the management of mental and substance abuse disorders should not only involve medication but psychosocial interventions also. We follow the multidisciplinary team approach in psychiatry. Psychiatric Social Work (PSW) is an integral part of the care, treatment, and rehabilitation of persons with mental illness along with other disciplines, e.g. psychiatry, psychology, and psychiatric nursing. Very often, in a multidisciplinary team, more than one professional is working or treating the same case. So, many a time there is a lack of clarity or connivance who will work on what issue and how to decide about a referral to another team member, particularly to psychiatric social worker. In this juncture, we have developed a screening tool, the 'Screening Tool for Assessment of Psychosocial Problems (STAPP)'. This paper illustrates the process of development of the same along with its reliability and validity.

MATERIAL AND METHODS

This study was done at the Department of Psychiatry, Government Medical College & Hospital, Chandigarh, India, which is a multispecialty, teaching, tertiary care hospital. The study was approved by the departmental committee of the faculties. For reliability testing, only those caregivers of the patients of mental illness were included for the study who had given informed consent.

Development of a valid and reliable tool involves several steps taking considerable time. The sequential steps involved in the development of STAPP used in the present study are represented in Figure 1.

Step 1: Review of literature

Using various database (PubMed, Google Scholar), Google search engine, as well as searching published articles and research studies in the books and periodicals, the authors could not find an instrument for assessment of psychosocial problems in persons with mental illness or their caregivers. Therefore, this instrument, STAPP was developed. The review helped in planning the items and content of the tool. Following the review, the blueprint was prepared.

Step 2: Preparation of blueprint

After the literature review, we have considered the 12 following areas- knowledge/awareness, medication/treatment compliance, availability of financial resources, social support, expressed emotion, emotional/physical/sexual abuse, legal issues, conflicts including property and family, employment,

accommodation, stigma, and activities of daily living (ADL) for STAPP. Each one had one question to be asked to patient/caregiver, assessed in four points (zero to three) scale.

Step 3: Development of the items

On the basis of the blueprint, 12 items were developed. All factors that contribute to the quality of the test items were taken into consideration. Further, the items were constructed in question format clearly in such a way that the patient or caregiver (two different sets for these two group of respondents) can answer without looking at the given options. The possible answer includes four probable options. While drafting STAPP, each word of the questions were checked twice to make it clear. Efforts were made to keep sensitivity towards psychological state of the respondents, bias free, and consideration was given to reading level of them. The authors critically reviewed the draft before sending it to the experts for content validity.

Step 4: Content validity

Content validity ratio (CVR) was calculated to determine the content validity by asking the viewpoints of the panel of experts. STAPP along with its blueprint (meaning, purpose, number of items, etc.) which was prepared before finalising the items and criteria checklist were submitted to the professional experts from the fields of psychiatry, psychology, psychiatric nursing, and PSW; all together, ten.[1] Recommendations were adopted on observing grammar, using appropriate and correct words, applying correct and proper order of words in items, appropriate scoring as prescribed and used method for content validity in the literature.[1,2] The rationale behind keeping a limited number of experts was that if the number of experts is more, possibility of disagreement on items is more; most of the research keeps this number at five to ten.[3]

The experts were requested to specify whether an item was necessary for operating a construct in a set of items or not. To this end, they were requested to score each item from one to three with a three-degree range of "not necessary, useful but not essential, essential" respectively. CVR varies between -1 and 1. The higher score indicates further agreement of members of the expert panel on the necessity of an item in an instrument. The formula by Lawshe[4] of $CVR = (N_e - N/2)/(N/2)$, in which the N_e is the number of panellists indicating "essential" and N is the total number of panellists. The numeric value of CVR was determined by Lawshe table.[4] In the present study, the number of panellists was ten; so, CVR bigger than 0.62 for a particular item in the instrument was accepted as shown in Table 1.

Content validity index (CVI) was also used which is the most widely used approach for content validity of instrument development.[1,5,6] Panel members were asked to rate the instrument items in terms of clarity to the construct underlying study as per the theoretical definitions of the construct itself and its dimensions on a four-point ordinal scale: one (not clear), two (item needs major revision), three (clear but need minor revision), four (very clear), and its relevancy in the similar manner (one [not relevant], two [somewhat relevant], three [quite relevant but need minor revision], and four [highly relevant]).[1]
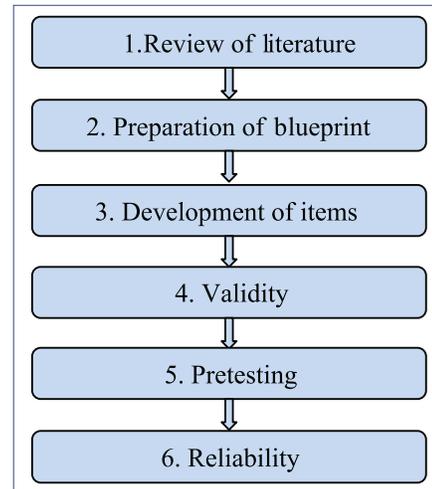

Figure 1: Sequential steps used in the development the Screening Tool for Assessment of Psychosocial Problems (STAPP)

Table 1: Calculating CVR at the first round

Items	N_e*	CVR	Interpretation
1	7	0.40	Eliminated
2	6	0.20	Eliminated
3	9	0.80	Remained
4	8	0.60	Eliminated
5	8	0.60	Eliminated
6	10	0.99	Remained
7	10	0.99	Remained
8	9	0.80	Remained
9	9	0.80	Remained
10	10	0.99	Remained
11	9	0.80	Remained
12	9	0.80	Remained

*Number of experts evaluated the item as "essential", CVR=Content Validity Ratio

CVI can be calculated both for item level (I-CVI) and scale-level (S-CVI). Item level, I-CVI was computed as the number of experts giving a rating of three or four to the relevancy of each item divided by the total number of experts as shown in Table 2. If I-CVI is higher than 59%, the item will be appropriate. If it is between 50 and 59%, it needs revision. If it is less than 50%, it is eliminated. I-CVI expresses the proportion of agreement on the relevancy of each item, which is between zero and one,[5,7] and S-CVI is defined as "the proportion of total items judged content valid"[5] or "the proportion of items on an instrument that achieved a rating of 3 or 4 by the content experts".[8] S-CVI can be calculated by two approaches: S-CVI/universal agreement (UA) and S-CVI/average number of experts (Ave).

Although CVI was calculated, kappa statistic was also calculated. Because, unlike CVI, it adjusts for chance agreement.[9] Chance agreement is an issue of concern while studying agreement indices among assessors, especially when we place four-point scoring within two relevant and not relevant classes.[10] In other words, kappa statistic is a

consensus index of inter-rater agreement that adjusts for chance agreement and is an important supplement to CVI because kappa provides information about the degree of agreement beyond chance.[10] Nevertheless, CVI is mostly used by researchers because it is simple for calculation, easy to understand, and provide information about each item, which can be used for modification or deletion of instrument items.[10,11]

To calculate the modified kappa statistic, the probability of chance agreement was first calculated for each item by the following formula: $PC=[N!/A! (N-A)!] \times 5N$. In this formula, N=number of experts in a panel and A=number of panellists who agree that the item is relevant. After calculating I-CVI for all instrument items, finally kappa was computed by entering the numerical values of probability of chance agreement (PC) and CVI of each item (I-CVI) in the formula: $K=(I\text{-}CVI - PC)/(1- PC)$ as shown in Table 3.

Table 2: Calculation of I-CVI at the first round

Items	Relevant (rating 3 or 4)	Not relevant (rating 1 or 2)	I-CVI	Interpretation
1	7	3	0.7	Appropriate
2	5	5	0.5	Needs revision
3	9	1	0.9	Appropriate
4	8	2	0.8	Appropriate
5	8	2	0.8	Appropriate
6	9	1	0.9	Appropriate
7	8	2	0.8	Appropriate
8	9	1	0.9	Appropriate
9	9	1	0.9	Appropriate
10	10	0	1	Appropriate
11	8	2	0.8	Appropriate
12	7	3	0.7	Appropriate

I-CVI=Item-Content Validity Index

Step 5: Pretesting

Pretesting of the tool is an essential step before establishing reliability to enhance its clarity and to ensure acceptance of the study by the participants, and also, to check the suitability of question-wording. "It is the trial administration of a newly developed instrument to identify flaws or assess the time requirement".[10] The tool was administered to 15 primary caregivers of patients and the same number of clinically stable patients in the inpatient and outpatient settings respectively of a tertiary care hospital. STAPP was found to be clear and understandable to the subjects. The average time taken to complete the tool was approximately ten to 15 minutes. The reliability of STAPP was established after the pretesting.

Step 6: Reliability

Reliability refers to "the degree of consistency or accuracy with which an instrument ensures the attribute it has been designed to measure. It refers to the ability of a questionnaire to consistently measure an attribute and how well the items fit together, conceptually". [12,13] To establish reliability, split-half method was adopted. All samples were divided into two groups based on the odd and even number of questions. Correlation of the test was done using the Karl Pearson correlation coefficient formula and the Spearman-Brown prophecy formula was used to compute the reliability of the whole test.

RESULTS

Content validity ratio

As shown in Table 1, few items have low CVR; so, those items were eliminated and new modified items were added in place of the eliminated items. In the second round, all 12 items were having CVR level higher than 0.90; so, all were accepted.

Content validity index

As shown in Table 2, one item had low I-CVIs which needed revision; that particular item was revised, and the suggestions

Table 3: Content validity index, modified kappa, and comprehensiveness of instrument dimensions and total instrument at the third round of judgement

Items	Number giving rating of 3 or 4 to relevancy of item	I-CVI	PC	K	Interpretation	Comprehensiveness of total instrument
1	7	0.7	2.134	1.265	Excellent	Proportion of consensus
2	5	0.7	3.0	1.25	Excellent	0.90
3	9	0.9	0.556	0.777	Good	
4	8	0.8	1.25	1.8	Excellent	
5	8	0.8	1.25	1.8	Excellent	
6	9	0.9	0.556	0.777	Good	
7	8	0.8	1.25	1.8	Excellent	
8	9	0.9	0.556	0.777	Good	
9	9	0.9	0.556	0.777	Good	
10	10	1	0.5	1.0	Excellent	
11	8	0.8	1.25	1.8	Excellent	
12	7	0.7	2.134	1.265	Excellent	

I-CVI=Item-level Content Validity Index, PC=probability of a chance occurrence, K=modified kappa

and remarks were incorporated as received from the experts. Filially, in the second round, all 12 items were appropriate as I-CVI were 0.70 or above. S-CVI was calculated as 0.97 which is acceptable.

Modified kappa

As shown in Table 3, evaluation criteria for kappa values above 0.74, between 0.60 and 0.74, and the ones between 0.40 and 0.59 are considered as excellent, good, and fair respectively. It was interpreted using guidelines described in Cicchetti and Sparrow.[14] All items of STAPP fall either on excellent or good level.

Reliability

For reliability, STAPP was administered to 100 primary caregivers of adult patients with mental illness diagnosed with ICD-10[15] criteria in the inpatient setting of a tertiary care hospital.

Sample characteristics

The sociodemographic profile of the caregivers is summarised in Table 4.

Score description for STAPP

Scores on the 12 items of the tool along with the total score on STAPP administered to 100 primary caregivers is shown in Table 5. Statistical analyses showed good internal consistency by Cronbach's alpha (r=0.86) and Spearman-Brown prophecy formula (r=0.83). Scores of all the 12 individual items were significantly correlated with the total score of the tool indicating acceptable sensitivity. A significant correlation was found among the scores of items indicating good construct validity of the tool.

Description of the final STAPP

At the final stage, 12 questions in 12 different areas (each one had one question) remained on STAPP, to be asked to patient/caregiver assessed in four points scale from zero to three.

Total score=36
- 25-36=Sever psychosocial problems, need immediate intervention
- 13-24=Moderate psychosocial problems, need urgent intervention
- Up to 12= Mild psychosocial problems, may need intervention.

Feasibility

In test administration, it took less than ten minutes and seems to be capturing the appropriate issues for which it intended to.

DISCUSSION

The aim of the study was to develop a valid and reliable screening tool for assessment of psychosocial problems which need psychosocial interventions for use in routine clinical practice. A 12-item tool was developed; obviously, it is not a tool which will screen psychosocial problems comprehensively as it is limited to only 12 different areas. It was meant to use in a

Table 4: Sociodemographic profile of caregivers of persons with mental illness

Variables	Variables category	Mean±SD/n (%)
Age (in years)		33.70±22
Sex	Male	67
	Female	33
Marital status	Single	49
	Married	48
	Divorced/separated	2
	Not known	1
Education	Illiterate	10
	Primary	7
	Middle	8
	Matric	19
	Inter/diploma	23
	Graduate	30
	Postgraduate	0
	Professional	3
Occupation	Professional	5
	Semi-professional	5
	Clerical/shop-owner/farmer	10
	Skilled/semi-skilled/unskilled	19
	Housewife	16
	Retired	3
	Student	15
	Unemployed	27
Family income (in rupees)	0-10,000	40
	10,001-20,000	16
	20,001-30,000	42
	30,001 and above	2
Religion	Hinduism	51
	Islam	0
	Sikhism	45
	Christianity	2
	Other	2
Family type	Nuclear	41
	Extended	3
	Joint	54
	Other	2
Locality	Urban	59
	Rural	37
	Semi-urban	4

SD=Standard Deviation

tertiary care setting where a multidisciplinary team approach is followed and more than one professional is working or treating the same case. Psychosocial interventions are delivered in the broad range of settings by various providers including psychologists, psychiatrists, social workers, counsellors/therapists, etc.[16] So, STAPP was developed to enhance clarity or connivance to decide about a referral to

Table 5: Score description for STAPP

Items	Area of psychosocial problems	Score				
		Mean	SD	1	2	3
1	Knowledge/ awareness	1.32	1.278	9	21	27
2	Medication/ treatment compliance	0.79	1.140	12	11	15
3	Availability of financial resources	1.44	0.556	50	47	0
4	Social support	1.44	0.538	52	46	0
5	Expressed emotion	2.73	0.664	6	9	83
6	Emotional/ physical/sexual abuse	2.79	0.640	6	3	89
7	Legal issues	2.43	0.655	6	42	51
8	Conflicts including property and family	1.37	1.220	29	9	30
9	Employment	2.69	0.734	1	14	80
10	Accommodation	1.55	0.575	37	59	0
11	Stigma	1.320	1.278	9	21	27
12	ADL	2.430	0.6552	6	42	51
13	Total	22.460	4.547	2	65	33

STAPP=Screening Tool for Assessment of Psychosocial Problems, SD=Standard Deviation, 1=Mild, 2=Moderate, 3=Severe, ADL=Activities of Daily Living

psychiatric social workers. Psychosocial problems are due to psychosocial factors which influence an affected person psychologically or socially. These are multidimensional constructs encompassing several domains.[17] Psychosocial factors may contribute to the development or aggravation of mental and physical disorders;[18] on the contrary, several psychiatric disorders may affect psychological and social aspects of individual's lives.[19] "Psychosocial interventions for mental health and substance use disorders are interpersonal or informational activities, techniques, or strategies that target biological, behavioural, cognitive, emotional, interpersonal, social, or environmental factors with the aim of improving health functioning and well-being".[16] Generally, in India, psychiatric social workers are preferred for more social interventions to address social factors and clinical psychologists mostly take up psychological factors. The score on STAPP by 100 primary caregivers revealed that all the caregivers felt some psychosocial problems in the person with mental illness; none reported that there was no problem. The tool considered severity of psychosocial problems by certain cut-offs which is accumulated score on various items on various psychosocial problems. One can argue that some time one psychosocial problem can be severe enough which need to be addressed.

All items were included by considering the more than 90% validity index. The tool was administered in 100 family members of persons with mental illness. Total score range found in the respondent was from six to 33 with mean of 22.46±4.55. This indicates the higher psychosocial problems which need intervention. This could be because the sample was taken from caregivers of the hospitalised patients where psychosocial problems are expected more.

The Spearman-Brown prophecy formula was found to be 0.83 in split-halves among odd and even items. Cronbach's alpha was found to be 0.86. As the reliability coefficients above 0.7 are considered satisfactory,[20] the results for STAPP indicated good internal consistency. The STAPP items were significantly correlated among themselves and with its total score which indicates good construct validity and sensitivity respectively. Convergent validity is the degree to which two measures of the construct are related[21] and which can be estimated by using the correlation coefficient.[22]

Limitations

Some limitations of content validity studies should be noted. The experts' feedback is subjective; thus, the study is subjected to bias that exists among the experts. If the content domain is not well-identified, this type of study does not necessarily identify that contents which have been omitted from the instrument. However, experts are asked to suggest other items for the instrument, which may help minimise this limitation.

Conclusion

In this paper, reliability and validity of a new screening tool, namely STAPP have been discussed which is constructed by the researchers using different methods. The satisfactory values of reliability and validity of the constructed tool mean that it is a reliable and valid screening tool for assessment of psychosocial problems.

REFERENCES

1. Davis LL. Instrument review: getting the most from a panel of experts. Appl Nurs Res. 1992;5:194-7.
2. Safikhani S, Sundaram M, Bao Y, Mulani P, Revicki DA. Qualitative assessment of the content validity of the Dermatology Life Quality Index in patients with moderate to severe psoriasis. J Dermatolog Treat. 2013;24:50-9.
3. Kerlinger FN, Lee HB. Foundations of behavioral research. Wadsworth Publishing; 1999.
4. Lawshe CH. A quantitative approach to content validity. Pers Psychol. 1975;28:563-75.
5. Lynn MR. Determination and quantification of content validity. Nurs Res. 1986;35:382-5.
6. Grant JS, Davis LL. Selection and use of content experts for instrument development. Res Nurs Health. 1997;20:269-74.
7. Waltz CF, Bausell RB. Nursing research: design, statistics, and computer analysis. Philadelphia: FA Davis FA; 1981.
8. Beck CT, Gable RK. Ensuring content validity: an illustration of the process. J Nurs Meas. 2001;9:201-15.
9. Wynd CA, Schmidt B, Schaefer MA. Two quantitative approaches for estimating content validity. West J Nurs Res. 2003;25:508-18.
10. Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. Res Nurs Health. 2007;30:459-67.
11. Polit DF, Beck CT. The content validity index: are you sure you know what's being reported? Critique and recommendations. Res Nurs Health. 2006;29:489-97.
12. Downing SM. Selected-response item formats in test development. In: Downing SM, Haladyna TM, editors. Handbook of test development. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers; 2006:287-301.

13. DeVon HA, Block ME, Moyle-Wright P, Ernst DM, Hayden SJ, Lazzara DJ, *et al*. A psychometric toolbox for testing validity and reliability. J Nurs Scholarsh. 2007;39:155-64.
14. Cicchetti DV, Sparrow SA. Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. Am J Ment Defic. 1981;86:127-37.
15. World Health Organization. The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines. Geneva: World Health Organization; 1992.
16. England MJ, Butler AS, Gonzalez ML. Psychosocial interventions for mental and substance use disorders: a framework for establishing evidence-based standards. Washington (DC): National Academy Press; 2015.
17. Suzuki SI, Takei Y. Psychosocial factors and traumatic events. In: Gellman MD, Turner JR, editors. Encyclopedia of behavioral medicine. New York, NY: Springer; 2013:1582-3.
18. Falagas ME, Zarkadoulia EA, Ioannidou EN, Peppas G, Christodoulou C, Rafailidis PI. The effect of psychosocial factors on breast cancer outcome: a systematic review. Breast Cancer Res. 2007;9:R44.
19. de Oliveira AM, Buchain PC, Vizzotto AD, Elkis H, Cordeiro Q. Psychosocial impact. In: Gellman MD, Turner JR, editors. Encyclopedia of behavioral medicine. New York, NY: Springer; 2013:1583-4.
20. Polit DF, Hungler BP. Nursing research: principles and methods. 6th ed. Philadelphia, PA: Lippincott Williams & Wilkins; 1999.
21. Trochim WMK. Convergent & discriminant validity. In: Research methods knowledge base [Internet]. 2006 Oct 20 [cited 2018 Nov 7]. Available from: https://socialresearchmethods.net/kb/convdisc.php.
22. Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychol Bull. 1959;56:81-105.