International Journal of Community Medicine and Public Health
Teja RNP et al. Int J Community Med Public Health. 2024 Nov;11(11):4296-4305
http://www.ijcmph.com
pISSN 2394-6032 | eISSN 2394-6040

DOI: https://dx.doi.org/10.18203/2394-6040.ijcmph20243289

Original Research Article

Development, validation, and reliability testing of the standard treatment workflow checklist

Rangisetti Naga Prudhvi Teja[1]*, Arjun Thimmiah[1], Kavita Kachroo[1], Syed Abdul Khader Moinudeen[1], Vanjavakam Sahithya[1], Jitendra Sharma [1], Ashoo Grover[2]

[1]Department of Health Research, Kalam Institute of Health Technology, Visakhapatanam, Andhra Pradesh, India
[2]Indian Council of Medical Research, New Delhi, Delhi, India

Received: 09 June 2024
Revised: 14 October 2024
Accepted: 18 October 2024

***Correspondence:**
Dr. Rangisetti Naga Prudhvi Teja,
E-mail: prudhvi.r@kiht.in

ABSTRACT

Background: Standard Treatment Workflows (STWs) are systematically developed documents designed to enhance clinical care and patient outcomes. Validating these workflows is essential for their adoption across diverse clinical settings. This study aimed to develop and validate a checklist for evaluating STWs.
Methods: The study was conducted in three phases. Initially, items were generated through a comprehensive literature review and expert consultations, followed by pre-testing (N=50). In the second phase, content validation (N=43) was performed using the Content Validity Index (CVI) and Content Validity Ratio (CVR) to assess the relevancy and essentiality of each item. Item-level (I-CVI) and scale-level (S-CVI) indexes were calculated. Items were accepted, modified, or rejected based on these scores. In the third phase, pilot testing (N=30) assessed the checklist's reliability and calculated Cronbach's alpha.
Results: The final checklist comprised 33 close-ended and one open-ended items. Each item was deemed relevant and essential, with I-CVIs ranging from 0.96 to 1.00 and S-CVI/UA and S-CVI/Ave values of 0.83 and 0.99, respectively. CVR values ranged from 0.85 to 1.00, with an average CVR of 0.97. Reliability testing yielded Cronbach's alpha values of 0.863 and 0.961 for the two parts of the checklist.
Conclusions: The developed checklist is valid and reliable, with satisfactory CVI, CVR, and Cronbach's alpha values. It can be used as a standard tool for assessing STW validation and effectively capturing essential aspects of STW evaluation.

Keywords: Content validation, Cronbach's alpha, Checklist development, Item-total correlation, Reliability, Standard treatment workflows

INTRODUCTION

Standard treatment workflows (STWs) are systematically developed documents designed to assist clinicians in providing better and more appropriate healthcare. There is standard treatment guidelines (STGs) and procedures from many organizations, however, they are dense, complex, and difficult to review. As a result, clinicians need some simple, basic, and easy-to-reference guidance materials that can be viewed in several ways. To meet this requirement, the Indian Council of Medical Research (ICMR) collaborated with the National Health Authority, the Government of India, and the World Health Organization (WHO) India Country Office to develop STWs. These workflows have been designed for common and serious diseases clinicians encounter across all healthcare system levels, from primary to tertiary care.[1]

The STWs are expected to bring a structured and accurate case management protocol in a concise format. They are essential for achieving uniform standards of care in the Indian public healthcare system, assisting with effective administration, and promoting proper utilization of drugs, healthcare services, and other diagnostics. These STWs, prepared by national specialists with significant expertise, evaluate the feasibility of implementation within India's healthcare system. Each STW is a one-page document with key indicative actions. As of July 30, 2024, STWs for 157 disease conditions across 28 specialties, compiled into four volumes, were available in both print and digital formats.

The primary purpose of STWs is to guide clinicians in the rational use of medicines, ultimately benefiting patients and working towards universal health coverage. In the health care system, STWs assist in formulating and budgeting services, becoming crucial for overseeing and approving processes in government-funded health insurance plans.[2]

Validation of STWs is paramount since no such concise, nationwide documents existed before their introduction by ICMR. Studies should ensure that STWs meet the desired quality and performance criteria, enabling intended users to utilize them effectively. While validated tools are available to ensure guideline transparency, quality, and completeness, directly using these tools to evaluate STWs is not feasible. These tools are designed to evaluate the detailed process from development to the final implementation of comprehensive practice guidelines, whereas STWs deliver critical information concisely, serving as quick reference guides.[3,4]

To address this evidence gap, we have developed a checklist for evaluating and appraising STWs in practice. Ensuring the validity and reliability of this new checklist is crucial to obtaining relevant information on STWs. Validity ensures that the collected data accurately covers the intended area of investigation, measuring what is intended to be measured. Reliability describes the "consistency and stability of measurements". A measurement tool is considered reliable if it provides consistent results under constant conditions.[5] Testing for both validity and reliability is essential to ensure that the developed checklist effectively evaluates and appraises STWs. Therefore, this study aims to develop, validate, and assess the reliability of the checklist developed to evaluate STWs.

METHODS

This methodology will be a part of a mixed-method explanatory study aimed at validating STWs and assessing their feasibility of implementation in diverse healthcare settings, both public and private, across the selected regions. This study was conducted in the first quarter of 2024. The study employs a cross-sectional design conducted in three phases: Phase One: Checklist Development. Phase Two: Validation of the checklist, focusing on content validity. Phase Three: Testing the reliability of the checklist, specifically its internal consistency. A detailed flow of the methodology is illustrated in Figure 1.

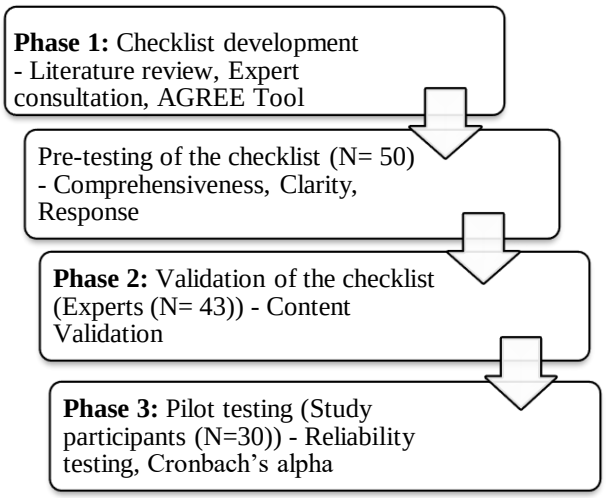

Figure 1: Methodology employed in the study.

Phase 1: checklist development

Content and item generation: The initial checklist items were developed by reviewing STW guidelines, conducting a comprehensive literature review on treatment workflow validation, consulting with experts, and examining healthcare quality assurance practices. The checklist is in a closed-ended format with three response options per item, where respondents select the option that best describes their answer. We developed the checklist following AGREE II guidelines for rigorous methodology and transparency. This included gathering and synthesizing evidence and continually refining items. By adhering to AGREE II, we ensured high quality, and reliability in evaluating treatment workflows.[4]

We generated around 34 items, focusing on adherence to evidence-based guidelines and regulatory compliance. We then created a checklist using Google Forms, ensuring accessibility via computers, smartphones, or tablets.

Pre-testing

Comprehensiveness, clarity, and response of the checklist: We pre-tested the checklist's comprehensiveness, clarity, and response options with the help of 50 experts. These aspects are defined as follows: 1) Comprehensiveness: Do the items cover all important aspects?, 2) Clarity: Are the items clear and easy to understand?, 3) Response Options: Are the response options appropriate and relevant to the constructs measured in the checklist?

Phase 2: checklist validation

A panel of healthcare professionals specialized in clinical practice, workflow management, and quality assurance conducted the content validation for the checklist. Domain experts from various departments were chosen based on their professional experience in relevant fields. They individually reviewed each item for essentiality and relevance, providing open-ended comments as needed. All experts gave informed consent via the Google Form before participating. A larger number of experts were involved to minimize agreement by chance. Each item was independently rated using a 4/3-point scale, guided by a scoring table in the Google Form to ensure consistency.

Content Validity Index (CVI): CVI method assesses "how well each item reflects the intended construct". Each item is rated by experts on a scale of 1 (not relevant) to 4 (very relevant). CVI is a common indicator of content validity, and it is calculated using Item-level-CVI (I-CVI) and Scale-level-CVI (S-CVI).

To compute I-CVI, we divide the number of experts who rate item 3 or 4 by the total number of experts. It involves dichotomizing ratings by combining 3 and 4 as relevant, and 1 and 2 as irrelevant, then calculating the agreement among experts. The I-CVI measures expert agreement on item relevance, with a range from zero to one, reflecting the degree of consensus among experts. Items with an I-CVI more than 0.79 are considered relevant, while those with values between 0.70 and 0.79 may require improvements, and items less than 0.70 are deleted.[6]

S-CVI assesses overall content validity. It can be calculated using "S-CVI/UA (Universal Agreement) or S-CVI/Ave (Average)". S-CVI/UA is the proportion of items evaluated as "very relevant" (I-CVI = 1) among all items, whereas S-CVI/Ave is the average of all I-CVIs. An S-CVI/UA ≥0.8 or an S-CVI/Ave ≥0.9 depicts strong content validity, ensuring the items accurately represent the intended content.

Kappa statistics: While CVI is widely used for content validity, Wynd et al suggested using the kappa statistic to account for chance agreement.[7] The kappa statistic measures agreement beyond chance using the following method:

$$K = \frac{(\text{ICVI} - \text{PC})}{(1 - PC)}$$

$$PC = \frac{(\text{N!})}{(A!\,(N - A))!} * (0.5^N)$$

This formula uses *P*C (probability of chance agreement), N (number of experts), and A (number of experts who consider an item relevant). Kappa values more than 0.74 are superior, 0.60 to 0.74 are good, and 0.40 to 0.59 are acceptable.

Content validity Ratio (CVR): "CVR measures the essentiality of each item, with scores of 1 (not essential), 2 (useful but not essential), and 3 (essential)". Ratings are dichotomized (combining 1 and 2 as non-essential), and agreement among experts is calculated accordingly. CVR is then computed using the following formula.

$$CVR = \frac{(\text{Ne}) - (N/2)}{(\text{N}/2)}$$

Where; Ne is the number of experts marked "essential" and N is the total number of experts. CVR scores range from -1 (complete disagreement) to +1 (complete agreement), with higher scores indicating greater agreement on an item's necessity. The CVR score is compared with the Lawshe Table to determine statistical significance. All the definitions are provided in the Table 1.

Table 1: The definition and formula of I-CVI, S-CVI/Ave, and S-CVI/UA.

CVI indices [8]	Definition	Formula
I-CVI (item-level content validity index)	The proportion of content experts giving the item a relevance rating of 3 or 4	I-CVI = (agreed item)/ (number of experts)
S-CVI/Ave (scale-level content validity index based on the average method)	The average of the I-CVI scores for all items on the scale or the average of proportion relevance judged by all experts. The proportion relevant is the average relevance rating by individual experts.	S-CVI/Ave = (sum of I-CVI scores)/ (number of items) (or) S-CVI/Ave = (sum of proportion relevance rating)/ (number of experts)
S-CVI/UA (scale-level content validity index based on the universal agreement method)	The proportion of items on the scale that achieve a relevance scale of 3 or 4 by all experts. Universal agreement (UA) score is given as 1 when the item achieved 100% experts in agreement, otherwise, the UA score is 0.	S-CVI/UA = (sum of I-CVI equals 1)/ (number of items)

Note: The definition and formula were based on the recommendations of Lynn et al[9], Davis et al[10], Polit et al[11]

Phase 3: reliability testing (pilot testing)

After validating the checklist, a pilot test was conducted with the final version of the checklist among 30 doctors providing services at tertiary care centers. A sample size of 30 was chosen based on literature review.[12] Internal consistency and reliability are evaluated using correlation analysis, such as Cronbach's alpha and item-total correlations. Item-total correlations below 0.3 indicate an insufficient correlation.[13] Cronbach's alpha evaluates response consistency on a scale of 0 to 1, with values closer to 1 indicating more reliability. The usually accepted minimum is 0.7.[14]

Data analysis

Data were exported from Google Forms, cleaned in Excel, and analyzed using SPSS version 17.0.

RESULTS

Phase 1

Comprehensiveness, clarity, and response results: Comprehensiveness: Of the 50 experts, 46 (90.2%) agreed that the items cover all important aspects. Clarity: Approximately 49 experts (96.1%) confirmed that the items are clear and easy to understand. Response Options: About 49 experts (96.1%) affirmed that the response options are appropriate and relevant to the constructs measured.

Phase 2

Socio-demographic details: A total of 43 experts participated in the study. Specifically, 3 experts evaluated only the CVI, 18 evaluated only the CVR, and 20 evaluated both the CVI and CVR, resulting in a total of 23 experts evaluating the CVI and 40 evaluating the CVR. The sample size for performing content validation was calculated by taking 10% of the original study's sample size (N=400). Descriptive statistics summarized the expert's demographics. The mean age was 41.7 (SD = 11.8) years, with an average of 12.8 (SD = 10.4) years of experience. Most experts were male (67.4%) and held MD and MS qualifications (65.1%). Common specialties included paediatrics (16.3%), general medicine (11.6%), and nephrology (9.3%) and the remaining are from various other departments. Many experts worked in tertiary care settings (81.4%) and were employed in the private sector (65.1%). Professors and Assistant Professors each comprised 20.9% of the group. Demographic details are presented in Table 2.

Table 2: Socio-demographic characteristics of the experts involved in content validation (n=43).

Variable	Mean (SD)	
Age in years	41.7 (11.8)	
Experience in years	12.8 (10.4)	
Variable	**Number**	**Percentage**
Gender		
Male	29	67.4
Female	14	32.6
Education qualification		
MD	28	65.1
MS	7	16.3
DM	3	7.0
MPH	2	4.7
MBBS DCP FRACP	1	2.3
MBBS MHA	1	2.3
MCH	1	2.3
Specialty		
Paediatrics	7	16.3
General medicine	5	11.6
Nephrology	4	9.3
Dermatology	3	7.0
Obstetrics and gynaecology	3	7.0
Orthopaedics	3	7.0
Gastroenterology	2	4.7
Ophthalmology	2	4.7
Pulmonology	2	4.7
Endocrinology	1	2.3
Neurology	1	2.3
Cardiology	1	2.3

Continued.

Variable	Mean (SD)	
Urology	1	2.3
Critical care associate	1	2.3
Emergency medicine	1	2.3
ENT	1	2.3
Oncology	1	2.3
Psychiatry	1	2.3
Radiology	1	2.3
Surgery	1	2.3
Surgical oncology	1	2.3
Current position		
Professor	9	20.9
Assistant professor	9	20.9
Consultant	8	18.6
Associate professor	5	11.6
HOD	4	9.3
Public health specialist	2	4.7
Senior resident	2	4.7
Chief surgical oncologist	1	2.3
Clinician	1	2.3
Director cloud physician	1	2.3
Medical superintendent	1	2.3
Practice setting		
Tertiary	35	81.4
Secondary	7	16.3
Primary	1	2.3
Sector		
Private	28	65.1
Government	15	34.9

Table 3: Calculation of the I-CVI for relevancy, CVR for essentiality for each item (n=34).

Item	Title of the item	I-CVI (relevancy)	Interpretation	CVR (essentiality)	Interpretation
Q1	The usage of terms like “workflows” or “recommendations” was mentioned in the title.	1.00	Relevant	1.00	Essential
Q2	The year of publication is stated in the STWs	1.00	Relevant	0.85	Essential
Q3	The aim and objective of the STW are clearly stated	1.00	Relevant	1.00	Essential
Q4	Is the "International Classification of Diseases (ICD)" mentioned/clearly stated in the STWs	0.96	Relevant	0.95	Essential
Q5	The STWs are successfully identifying the purpose such as screening, diagnosis, therapy, administration, and prevention of the clinical conditions	1.00	Relevant	1.00	Essential
Q6	The summary of the recommendations is mentioned in the STWs	1.00	Relevant	1.00	Essential
Q7	The abbreviations and acronyms are listed in the STWs	1.00	Relevant	1.00	Essential
Q8	Contact lists of corresponding authors/ developers are made accessible in the STWs	0.96	Relevant	0.95	Essential
Q9	The background of the disease/conditions is stated in the STWs	0.96	Relevant	1.00	Essential
Q10	The STWs have special consideration for the different patient subgroups.	1.00	Relevant	1.00	Essential
Q11	Details of contributors involved in developing the STWs contain their names, titles, roles, and	0.96	Relevant	0.75	Essential

Continued.

Item	Title of the item	I-CVI (relevancy)	Interpretation	CVR (essentiality)	Interpretation
	affiliations.				
Q12	The key questions in STWs have followed the PICO	1.00	Relevant	0.95	Essential
Q13	Appropriate methodology has been used for developing STWs	0.96	Relevant	1.00	Essential
Q14	Clear, precise, and actionable recommendations regarding the clinical conditions are clearly stated	1.00	Relevant	1.00	Essential
Q15	The STWs indicate the strength of recommendations and the certainty of the supporting evidence	1.00	Relevant	1.00	Essential
Q16	The cost and resource implications of the suggested interventions are judiciously considered	1.00	Relevant	0.90	Essential
Q17	The conflicts of interest are evaluated and managed and how users of the STW can access the declarations is stated in the STWs	1.00	Relevant	1.00	Essential
Q18	Any significant gaps in the evidence can be identified through STWs.	1.00	Relevant	0.90	Essential
Q19	Any suggestions for future research are identifiable through STWs.	1.00	Relevant	1.00	Essential
Q20	The recommendations are clear and specific.	1.00	Relevant	1.00	Essential
Q21	The multiple options for management of the health conditions or health issues are presented in STWs	1.00	Relevant	1.00	Essential
Q22	All the key recommendations suggested in the STWs are easily identifiable	1.00	Relevant	1.00	Essential
Q23	The STWs provide advice and/or tools on how the recommendations can be put into practice	1.00	Relevant	1.00	Essential
Q24	Consistency and quality of service and healthcare	1.00	Relevant	1.00	Essential
Q25	Reduction in performance variations	1.00	Relevant	0.95	Essential
Q26	Reduction in procedural and protocol mix-ups	1.00	Relevant	1.00	Essential
Q27	Reduction in misinterpretation or miscommunication of information	1.00	Relevant	1.00	Essential
Q28	Effective management of departmental systems	1.00	Relevant	0.95	Essential
Q29	Helps in cultivating transparent functions in clinical practice	1.00	Relevant	1.00	Essential
Q30	Implements error prevention strategies	1.00	Relevant	1.00	Essential
Q31	Appropriate records help to measure the iatrogenic complications	1.00	Relevant	1.00	Essential
Q32	Facilitate corrective actions in practice	1.00	Relevant	0.95	Essential
Q33	Transfer knowledge and skill	1.00	Relevant	1.00	Essential
Q34	Kindly provide your critical comments in the form of suggestion	1.00	Relevant	1.00	Essential
S -CVI/Ave		Average of I-CVIs = 0.99			
S -CVI/UA		29/34 = 0.85			
CVR/Ave		0.98			

I-CVI= item-level content validity index; CVR= content validity ratio; S-CVI/Ave= scale-level content validity index/average, S-CVI/UA = scale-level content validity index/ universal agreement, CVR/Ave = Average content validity ratio

Content validity results

Content validity (CVI and CVR) was calculated for all 34 items. All the results are presented in Table 3.

I-CVI results: The I-CVI analysis, given in Table 1, shows that all items were deemed important, with I-CVI values ranging from 0.96 to 1.00. Specifically, 29 items had an I-CVI of 1.00, whereas five had an I-CVI of 0.96. All I-CVI values exceeded the 0.79 criterion, hence all

items were considered relevant, and none were eliminated.

S-CVI results: This checklist has high content validity, as indicated by the S-CVI/UA and S-CVI/Ave metrics. The S-CVI/UA is 0.85, computed as the ratio of items having an I-CVI of 1.00 (29 items) to the total of 34 items, which is greater than the recommended threshold of 0.8. The S-CVI/Ave is extremely high at 0.99, calculated by dividing the sum of all I-CVIs (33.7) by the total number of items, which is above the 0.9 criterion. The Kappa values for the items ranged from 0.95 to 1.00, falling into the "excellent" category based on established benchmarks.

CVR results: The CVR scores ranged between 0.75 and 1.00, with every single item classified as "essential." Specifically, 24 items (70.6%) had a CVR of 1.00, nine items (26.5%) had a CVR of 0.85 to 0.95, and only one item (2.9%) had a CVR of 0.75. The average CVR for all items was 0.98. All CVR values exceeded the 0.29 threshold required for a group of 40 experts, hence every item is considered essential.[15]

Checklist refinement

Given the high CVI, CVR, and Kappa statistics values, the checklist was not refined, and a reliability assessment was performed in the next phase.

Phase 3

Socio-demographics: The pilot test involved 30 study participants. The mean age (SD) of the participants was 30 (35.1) years. Of these participants, 16 (53%) were male and 14 (47%) were female. Among them, 11 (37%) reviewed STWs for dermatology, 5 (17%) reviewed STWs for psychiatry, and the remaining participants reviewed STWs for various other departments.

Reliability: The checklist was divided into two parts: part one consisted of 23 items with 3 response options, and part two comprised a 10-item Likert scale. Reliability testing was conducted separately for each part.

Table 4: Calculation of the inter item correlation for part 1 of the checklist (n=23).

Items	Corrected item-total correlation	Cronbach's alpha if item deleted
1. The usage of terms like "workflows" or "recommendations" were mentioned in the STW title	0.138	0.862
2. The year of publication is stated in the STWs	0.730	0.847
3. The aim and objective of the STW are clearly stated	0.207	0.862
4. Does the "International Classification of Diseases (ICD)" is mentioned/clearly stated in the STWs	0.492	.853
5. The STWs are successfully identifying the purpose such as screening, diagnosis, therapy, administration, and prevention of the clinical conditions	0.490	0.853
6. The summary of the recommendations is mentioned in the STWs	0.598	0.854
7. The abbreviations and acronyms are listed in the STWs	-0.018	0.865
8. The contact list of corresponding authors/developers is made accessible on the STWs (website)	0.593	0.849
9. The background of the disease/conditions is stated in the STWs	0.353	0.858
10. The STWs have special consideration for the different patient subgroups	0.296	0.859
11. Details of contributors involved in developing the STWs contain their names, title, roles, and affiliations	0.449	0.855
12. The key questions in STWs have followed PICO (Population, Intervention, Comparator and Outcome) format	0.666	0.846
13. Appropriate methodology has been used for developing STWs	0.582	0.850
14. Clear, precise, and actionable recommendations regarding the clinical conditions are clearly stated	0.240	0.861
15. The STWs indicate the strength of recommendations and the certainty of the supporting evidence	0.697	0.844
16. Cost and resource implications of the suggested interventions are judiciously considered (availability of laboratory investigations at the primary level as per Indian Public Health Standards)	0.677	0.844
17. The conflicts of interest are evaluated and managed and how users of the STW can access the declarations are stated in the STWs	0.461	0.858
18. Any significant gaps in the evidence can be identified through STWs	0.616	0.848
19. Any suggestions for future research are identifiable through STWs	0.443	0.855
20. The recommendations are clear and specific	0.537	0.851

Continued.

Items	Corrected item-total correlation	Cronbach's alpha if item deleted
21. The multiple options for management of the health conditions or health issues are presented in STWs	0.034	.869
22. All the key recommendations suggested in the STWs are easily identifiable (highlighted)	0.240	0.861
23. The STWs provide advice and/or tools on how the recommendations can be put into practice	0.191	0.862

Table 5: Calculation of the inter item correlation for part 2 of the checklist (n=10).

Items (N= 10)	Corrected item-total correlation	Cronbach's alpha if item deleted
1. Consistency and quality of service and healthcare	0.874	0.957
2. Reduction in performance variations	0.716	.963
3. Reduction in procedural and protocol mix-ups	0.806	0.960
4. Reduction in misinterpretation or miscommunication of information0.	0.770	0.962
5. Effective management of departmental systems	0.865	0.958
6. Helps in cultivating transparent functions in clinical practice	0.887	0.957
7. Implements error prevention strategies	0.896	0.957
8. Appropriate records help to measure the iatrogenic Complications	0.770	0.961
9. Facilitate corrective actions in practice	0.887	0.957
10. Transfer knowledge and skill	0.893	0.957

Part 1: Part 1 of the scale indicated strong internal consistency, with a Cronbach's alpha value of 0.861, indicating a reliable and consistent measurement. Most items' corrected item-total correlations were above 0.3, showing they contributed appropriately to the scale. Although a few items (e.g., items 1, 7, and 21) had lower item-total correlations, their impact on overall reliability was minimal, and they were not removed because experts identified them as important. When an item is deleted, the Cronbach's alpha values range from 0.844 to 0.869. None of the items, if removed, significantly enhance the overall alpha. This suggests that all items are reasonably contributing to the reliability of the scale. Results are presented in Table 4.

Part 2: The 10-item Likert scale exhibited strong internal consistency, with Cronbach's alpha at 0.963, indicating high reliability in measuring the intended construct. All items showed strong item-total correlations, and removing any item would not significantly affect the overall reliability. Results are presented in Table 5.

DISCUSSION

The development and validation of a checklist for evaluating STWs were methodically executed to ensure its reliability and relevance across diverse clinical settings. This study demonstrated the efficacy of the developed checklist through rigorous phases of literature review, pre-testing, content validation, and reliability testing. The initial phase focused on generating relevant items through an extensive literature review and expert consultations. This approach ensured that the checklist covered all critical aspects of STWs, providing a robust foundation for subsequent validation steps.

The content validation process involved 43 experts who assessed the items for relevance and essentiality. CVR and CVI were pivotal in determining the relevancy and essentiality of each item. The high I-CVI values (0.96 to 1.00) and CVR values (above 0.85) reflect a consensus among experts on the essentiality of the items. The S-CVI scores for “universal agreement (S-CVI/UA) and average (S-CVI/Ave)” are 0.83 and 0.99, respectively. These indicators show a high level of agreement among experts, which validates the checklist's content. Cronbach's alpha for the two parts of the checklist were 0.863 and 0.961, respectively, based on reliability testing using correlation analysis involving 30 participants. These results signify high internal consistency, confirming that the checklist items are reliably measuring the intended aspects of STW validation.

Some studies complement one another by focusing on distinct areas of clinical practice and validation, such as implementing guidelines and evaluating new technologies. While some highlight the actual application of treatment procedures, others stress the significance of strong validation methodologies to ensure reliability and effectiveness. Similar studies have undertaken validation processes for clinical tools. For instance, Zadeh et al conducted a content validity study for a nursing assessment tool, achieving high I-CVI values and demonstrating the tool's reliability.[16] Similarly, a study on

scale development by Grover et al emphasized the systematic approach in formulating standard treatment guidelines, ensuring their applicability across various healthcare levels.[2] Sharma et al undertook a study to analyze the adoption of STG in India and identified various facilitators and barriers. The findings highlighted the importance of accessible medical information and the practical challenges faced by clinicians in implementing these guidelines.[17]

Bano et al conducted research comparing different diagnostic and treatment approaches, aiming to identify new gold standards for translational research. The study emphasized the need for proper validations to ensure the effectiveness of new methodologies.[18] Another study evaluated the accuracy and trueness of digital impressions compared to traditional gypsum checks. This study provided insights into the clinical application of digital technologies and their reliability in dental practice.[19]

These studies complement each other by addressing different aspects of clinical practice and validation, from guideline implementation to the evaluation of new technologies. While some focus on the practical adoption of treatment workflows, others emphasize the importance of robust validation methods to ensure reliability and effectiveness. Hence, the validation of STWs in India is crucial for ensuring quality healthcare delivery and achieving universal health coverage. The development of comprehensive STGs in India has been a significant endeavor, facing challenges such as a lack of common STG, a weak understanding of the concept, and managing consensus between specialists and generalists.[2]

Conformance-checking methodologies using neural networks have been proposed to ensure adherence to standard treatment plans extracted from reputable healthcare websites. Reviewing available guidelines globally has highlighted the need for coordinated efforts, standardized development processes, and a designated authority for STG development to ensure evidence-based, rigorously developed guidelines with wide accessibility and acceptance.[20] Validating STWs in India and aligning with similar articles, the healthcare system can enhance implementing evidence-based methods will enhance patient outcomes and aid in achieving universal health coverage. This study is crucial because it addresses a gap by developing a valid checklist for the validation of STWs. However, it has some limitations. We intend to do the Principal Component Analysis and exploratory factor analysis as part of improving the checklist with the continuous dynamic data being obtained from clinicians. Since the study involves ever-evolving STWs it becomes necessary to improve the quality of the checklist alongside.

CONCLUSION

The study's findings indicate that the developed checklist has strong content validity and reliability, as evidenced by the I-CVI, S-CVI, CVR, and Cronbach's alpha scores. As of our knowledge, this is the first checklist developed in India to assess the validity of STWs. We anticipate that the developed checklist will serve as a standard tool for evaluating the validity of STWs, as it has been illustrated with high validity and reliability. We utilized this checklist to conduct our further study on the validity of the STWs currently in practice.

ACKNOWLEDGEMENTS

We would like to thank all the participants who participated in the study.

Funding: This study was funded by the ICMR
Conflict of interest: None declared
Ethical approval: The study was approved by the AMTZ IC SCR with the clearance number REGN. #ECR/1702/INST/AP/2022

REFERENCES

1. STWs. Indian Council of Medical Research | Government of India. Available at: https://www.icmr.gov.in/stws. Accessed on 15 May 2024.
2. Grover A, Bhargava B, Srivastava S, Sharma LK, Cherian JJ, Tandon N, et al. Developing Standard Treatment Workflows-way to universal healthcare in India. Front Public Health. 2023;11:1178160.
3. EQUATOR Network. A Reporting Tool for Practice Guidelines in Health Care: The RIGHT Statement. Available at: https://www.equator-network.org/reporting-guidelines/right-statement/. Accessed 12 July 2024.
4. EQUATOR Network. The AGREE Reporting Checklist: a tool to improve reporting of clinical practice guidelines. Available at: https://www.equator-network.org/reporting-guidelines/the-agree-reporting-checklist-a-tool-to-improve-reporting-of-clinical-practice-guidelines/. Accessed 24 June 2024.
5. Taherdoost H. Validity and reliability of the research instrument; how to test the validation of a questionnaire/survey in a research. Int J Acad Res Manage. 2016;5:28-36.
6. Shi J, Mo X, Sun Z. Content validity index in scale development. Zhong Nan Da Xue Xue Bao Yi Xue Ban. 2012;37(2):152-5.
7. Wynd C, Schmidt B, Schaefer M. Two quantitative approaches for estimating content validity. Western J Nurs Res. 2003;25:508-18.
8. Yusoff MS. ABC of content validation and content validity index calculation. Educat Medi J. 2019;11(2):49-54.
9. Lynn MR. Determination and quantification of content validity. Nurs Res. 1986;35(6):382-5.
10. Davis LL. Instrument review: Getting the most from a panel of experts. Applied Nursing Research. 1992;5(4):194-7.

11. Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. Res Nurs Health. 200;30(4):459-67.
12. Bujang MA, Omar ED, Foo DHP, Hon YK. Sample size determination for conducting a pilot study to assess reliability of a questionnaire. Restor Dent Endod. 2024;49(1):e3.
13. Jansen M, Doornebosch AJ, de Waal MW, Wattel EM, Visser D, Spek B, et al. Psychometrics of the observational scales of the Utrecht Scale for Evaluation of Rehabilitation (USER): Content and structural validity, internal consistency and reliability. Arch Gerontol Geriatr. 2021;97:104509.
14. Gottems LBD, Carvalho EMPD, Guilhem D, Pires MRGM. Good practices in normal childbirth: reliability analysis of an instrument by Cronbach's Alpha 1. Rev Lat Am Enfermagem. 2018;26:e3000.
15. Ayre C, Scally A. Critical Values for Lawshe's content validity ratio. Measurement and Evaluation in Counseling and Development. 2013;47:79-86.
16. Zamanzadeh V, Ghahramanian A, Rassouli M, Abbaszadeh A, Alavi-Majd H, Nikanfar AR. Design and implementation content validity study: development of an instrument for measuring patient-centered communication. J Caring Sci. 2015;4(2):165-78.
17. Sharma S, Pandit A, Tabassum F. Potential facilitators and barriers to adopting standard treatment guidelines in clinical practice: An Indian context. Int J Heal Care Quality Assur. 2017 ;30(3):285-98.
18. Bano R, Gupta S, Shekhar C. Translational research in biomedical sciences in India: Challenges, observations & national perspectives. Ind Med Res. 2020;152(4):335.
19. Barile G, de Rosa G, Papadia P, Muci G, Capodiferro S, Corsalini M. Clinical comparative study for validation of digital impression reliability with the gypsum check: a simple and fast way to evaluate the trueness and accuracy of implant-supported rehabilitation. Appli Sci. 2023;13(13):7358.
20. Lahariya C, Sharma S, Agnani M, de Graeve H, Srivastava JN, Bekedam H. Attributes of standard treatment guidelines in clinical settings and public health facilities in India. Indian J Community Med. 2022;47(3):336-42.

Cite this article as: Teja RNP, Thimmiah A, Kachroo K, Moinudeen SAK, Sahithya V, Sharma J, et al. Development, validation, and reliability testing of the standard treatment workflow checklist. Int J Community Med Public Health 2024;11:4296-305.