



Contents lists available at ScienceDirect

Journal of Ayurveda and Integrative Medicine

journal homepage: <http://elsevier.com/locate/jaim>

Review Article

Methodology for developing and evaluating diagnostic tools in Ayurveda – A review

Mukesh Edavalath ^{a, *}, Benil P. Bharathan ^b^a Department of Roganidana, VPSV Ayurveda College, Kottakkal, Kerala, India^b Department of Agadatantra, VPSV Ayurveda College, Kottakkal, Kerala, India

ARTICLE INFO

Article history:

Received 22 September 2020

Received in revised form

11 January 2021

Accepted 15 January 2021

Available online 5 March 2021

Keywords:

Ayurveda diagnosis

Ayurvedic diagnostic tool development

Diagnostic accuracy

Diagnostic test assessment

Gold standard

ABSTRACT

Ayurveda has a holistic and person-centric approach towards health and disease, which in turn necessitates consideration of several factors in the process of a diagnostic workup. This concept of personalised diagnosis brings about a high level of variability among the clinicians with respect to their assessment methods and disease diagnosis. Developing and validating diagnostic tools for diseases enumerated in the Ayurvedic classical textbooks can help in standardising the clinical approach, even when attempting to arrive at a patient specific diagnosis. However, diagnostic research is a very less explored area in Ayurveda and there are no established standards for developing and evaluating diagnostic tools. This paper reviews the methodology for the development and validation of diagnostic tools, available in published literature and proposes to integrate this in the field of Ayurveda. The search was conducted on online databases including PubMed, Science Direct, Scopus, and Google scholar, with keywords - ayurvedic diagnosis, diagnostic tool development, validity, reliability, and diagnostic test assessment. The articles were screened based on their comprehensiveness, relevance, and feasibility, and the methodology elaborated in the selected articles was organized into a framework that can be adopted in Ayurveda. We have also tried to examine the methodological challenges of integrating the fundamentals of ayurvedic diagnosis within the current methods of diagnostic research and explored possible solutions. The proposed tool development process involves both qualitative and quantitative components, which may be carried out in three phases that include setting the diagnostic criteria, tool development and validation, and diagnostic test assessment.

© 2021 The Authors. Published by Elsevier B.V. on behalf of Institute of Transdisciplinary Health Sciences and Technology and World Ayurveda Foundation. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Diagnosis is a fundamental component in clinical practice, with significant implications on patient care, research, and health policy. A clinician uses the method of reasoning, based on the scientific knowledge of diseases and their manifestations, for arriving at a diagnosis. Yet, in many instances, this process of reasoning can lead to pitfalls, especially when the factors at hand are more complex, including the variance in individual approach, population characteristics, the prevalence of the disease, and even the facilities at hand. In Ayurveda, owing to its holistic and patient-centric approach, the diagnostic process involves the assessment of

several subjective and objective parameters pertaining to the disease as well as the patient [1]. Thus, even in patients with identical modern diagnosis, the ayurvedic therapeutic indicators may vary, generating a personalised diagnosis. This scheme of a workup can bring about discrepancies among clinicians, with regards to the application of assessment methods and diagnostic workup, eventually contributing to a poor interobserver agreement [2,3]. However, in clinical practice, since ayurvedic medicines are formulated based on their action on respective *dosha* (body humor), precise disease nomenclature may not have much implication on the treatment outcome. But in research, this poses a disadvantage, as reproducibility cannot be achieved if there is a disagreement in the diagnosis [4,5]. Currently, most of the ayurvedic clinical trials employ generic tools that are based on references from classical textbooks, albeit not validated scientifically. Further, since the pathological concepts of Ayurveda are based on the functional derangements of *dosha*, there are also no established gold

* Corresponding author.

E-mail: mukeshayur@gmail.com

Peer review under responsibility of Transdisciplinary University, Bangalore.

standards like histopathology, for validating these tools. Hence, there is a need for developing and validating diagnostic tools for diseases enumerated in Ayurveda, to standardise the clinical approach, even when attempting for a patient-centric, individualised diagnosis. However, diagnostic research is a very less explored area in Ayurveda and there are no widely accepted standards for developing and evaluating diagnostic tools either. This paper reviews the current methodology for the development and validation of diagnostic tools and proposes to incorporate this in the field of Ayurveda. The search for articles was conducted on online databases including PubMed, Science Direct, Scopus, and Google scholar, with keywords - ayurvedic diagnosis, diagnostic tool development, validation of diagnostic tool, validity, reliability, and diagnostic test assessment. A total of 452 articles were listed, of which 59 were screened based on their comprehensiveness in the elaboration of methodology, relevance, and feasibility. The process of developing and validating health assessment tools was reviewed and organised into a framework that could be adopted in Ayurveda. The challenges of integrating the fundamentals of ayurvedic diagnosis within the concepts of modern diagnostic research are also highlighted with their possible solutions (Fig. 1).

2. Current concepts of diagnosis

Diagnosis can be considered as both a process and a classification scheme, which in turn is described as a “pre-existing set of categories agreed upon by the medical profession to designate a specific condition” [6]. The diagnosis of a disease may be confirmed with the help of history and physical examination, a specific test or investigation, or a combination of the above said factors, which are collectively known as diagnostic criteria. A good diagnostic tool or test should enable a clinician to perfectly distinguish between patients with and without a certain condition [7]. However, not all diagnoses can be confirmed with the help of an investigation like imaging or histology, in case of which, an algorithmic method with a sequential approach may help arrive at the diagnosis. There are also composite health measurement scales that are either clinician

elicited or patient-reported, used for diagnosis or assessment of treatment outcomes [8].

3. Classification of a modern diagnostic tool/test

Based on the purpose for which they are used, diagnostic measures can be broadly divided into 3 categories: discrimination, prediction, and evaluation [9]. The discriminative ability of a measure is vital to differentiate between patient groups or identify meaningful differences in patients' conditions. A predictive measure, as the name implies, is used to predict the outcome or prognosis, so that clinicians can select appropriate treatment goals and strategies. An evaluation tool is useful in detecting the longitudinal change in an individual or group, after the initiation of therapy. There is another classification based on the specific role of the test or tool, as follows [10].

(1) Screening – to identify the disease in an apparently healthy person; (2) Triage – a rapid test that can be applied with minimal false positives; (3) Confirmation/exclusion – confirm or exclude a disease; (4) Monitoring – to assess the efficacy of treatment; (5) Prognosis – assessment of outcome or disease progression.

Depending upon who is administering it, the tool can be classified under two [11].

(1) Proxy or physician-administered tool – these are tools administered by a physician or healthcare worker, aimed at diagnosing or classifying disease, and includes identification of subjective manifestations and objective signs in the patient; (2) Self-administered tool – These tools are used in certain disorders that do not have objective signs, as in headache or other subjective manifestations, where the patients themselves report their symptoms. Also, tools intended at monitoring therapeutic responses or other evaluations like the quality of life falls under this category.

4. Ayurvedic diagnostic process – concepts, challenges, and solutions

Even though working on the same patient, there are significant differences between Ayurveda and modern biomedicine, with regards to the understanding of diseases and methods of assessment. While the former relies on the concept of *dosha* imbalances elicited by signs and symptoms, the latter banks on the changes in the molecular mechanisms, which in turn are identified by meticulous investigations. Moreover, for personalizing the therapy, ayurvedic diagnostics rely on an algorithmic approach, that necessitates the assessment of various subjective and objective parameters relating to the patient (*Rogi pareeksha*) as well as the manifested disease (*Roga pareeksha*) (Table 1) [12].

The assessment of *shatkriyakala* (therapeutic windows) is also unique to Ayurveda, which helps in identifying the pathogenesis, even at a very early stage [13]. It is also to be noted that the final diagnosis does not merely provide a disease name, rather puts forward multiple therapeutic indicators, enabling the physician to arrive at an individualized diagnosis. This eventually prompts variations in the prescription, even between patients with an identical modern diagnosis, and also renders a one-on-one correlation of diseases entailed in Ayurveda and biomedicine, an unrewarding exercise. Such an elaborate scheme of assessment necessitates the development of diagnostic measures in the form of questionnaires or composite health measurement tools, that are based on disease descriptions elaborated in the classical textbooks. After devising the tools, validation should be carried out systematically, including the assessment of psychometric properties and diagnostic accuracy.

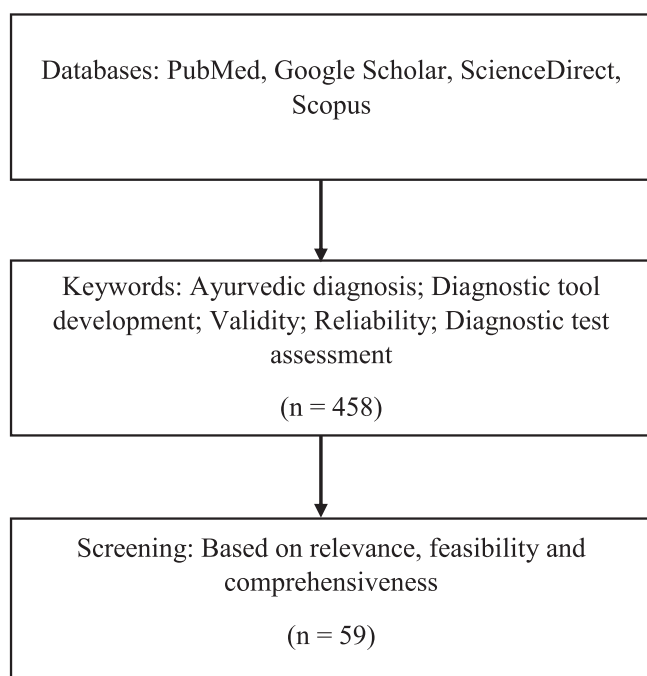


Fig 1. Flow chart of the review process.

5. Scope of diagnostic tools in Ayurveda

Ayurvedic classical textbooks comprise extensive references regarding *lakshanas* (signs and symptoms) of each disease, which can be employed in developing tools for clinical purposes. For example, the *Poorvaroopas* [14] (premonitory signs) can predict future diseases whereas *asadhya lakshana* (signs of bad prognosis) could predict an unfavourable outcome of the respective disease. Certain *lakshanas* like *sama* and *nirama* (two different stages of a disease), *rogamukta lakshanas* (features of remission), etc indicate the therapeutic response.

Accordingly, tools falling under the following domains can be developed in Ayurveda.

Diagnostic tools: These are tools aimed at predicting, diagnosing, or grading specific disorders enumerated in classical textbooks; **Classification tools:** After diagnosing the disease, it has to be classified based on parameters like *dosha* predominance or stage of the disease. Or, when an apt diagnosis is not available, classification can be made based on therapeutic indicators like *dosha* and *dhatu* (body tissues) involved, *srotas* (body channels), *Prakriti*, *Agni*, etc.; **Monitoring or evaluation tools:** In most instances, the diagnostic tool itself can serve the purpose of evaluation, provided a grading system is also incorporated. However, in some disorders, a separate tool might be necessary for evaluating the therapeutic response; **Prognostic tool:** Ayurvedic literature is abundant with signs and symptoms which indicate good as well as bad prognosis, which in turn, can be used to make predictions regarding the probable therapeutic outcomes.

As mentioned earlier, these tools can either be devised in the form of a physician administered one, with an added section to assess patient reported variables or in a self-administered form, especially for diagnosing diseases characterized by subjective manifestations alone.

6. Qualities of a good tool

Any measuring instrument which is in the form of a questionnaire, statement, or checklist, should possess certain qualities to be adopted into clinical practice, as described below [15,16].

Adequate for the problem intended to be measured (content validity); Reflect underlying theory or concept to be measured (construct validity); Reliability and precision, so that the measurements are consistent; Feasibility – simple and acceptable to patients and physicians; Sensitivity to change – capable of measuring change through time.

Of the above, the instrument has to be evaluated extensively for its psychometric properties, including validity and reliability, during the development stage itself.

Table 1
Assessment parameters in Ayurvedic diagnosis.

Rogi pareeksha	Roga pareeksha	External factors
<i>Prakriti, Sara, Samhanana, Satwa Satmya, Pramana, Agni, Vyayamasakti, Vaya.</i>	<i>Dosha, Dushya, Rogamarga, Nidana, Samprapti, Poorvarupa, Rupa, Upadrava, Upasaya anupasaya, Roga avastha – Sama/Nirama, Nava/Purana</i>	<i>Kala, Desa</i>
<i>Prakriti</i> (body constitution), <i>Sara</i> (excellence in dhatu), <i>Samhanana</i> (body composition), <i>Satwa</i> (mental strength), <i>Satmya</i> (preferences), <i>Pramana</i> (anthropometry), <i>Agni</i> (digestive power), <i>Vyayamasakti</i> (exercise tolerance), <i>Vaya</i> (age).	<i>Dosha</i> (body humor), <i>Dushya</i> (vitiated tissues), <i>Rogamarga</i> (disease pathway), <i>Nidana</i> (etiology), <i>Samprapti</i> (pathogenesis), <i>Poorvarupa</i> (premonitory signs), <i>Rupa</i> (clinical manifestations), <i>Upadrava</i> (complications), <i>Upasaya anupasaya</i> (explorative therapy), <i>Roga avastha</i> (stage of disease), <i>Sama</i> (with ama/toxic metabolite), <i>Nirama</i> (without ama), <i>Nava</i> (recent onset), <i>Purana</i> (chronic), <i>Kala</i> (season), <i>Desa</i> (place of residence).	

7. Phases of tool development and their implications in Ayurveda

The proposed framework of diagnostic tool development consists of both qualitative and quantitative components, which can be carried out under three phases.

Preliminary phase – defining or setting the diagnostic/classification criteria; The second phase – tool development and validation; Third phase – diagnostic tool assessment.

7.1. Preliminary phase – defining the diagnostic/classification criteria

Even though used synonymously in various instances, there is a difference between diagnostic criteria and a diagnostic tool [17], especially in the case of Ayurveda. The criteria provide the list of symptoms and signs essential for diagnosing a specific disease [18], whereas the tool aids in eliciting those in the given patient. Even though the criteria can serve as a tool in many of the conditions, this may not be practical always. E.g. – in the case of the disease *amavata*, *lakshanas* enumerated include *angamarda* (body ache), *aruchi* (anorexia), *trishna* (thirst), *alasya* (lassitude), *gourava* (heaviness), *jwara* (fever), *apaka* (indigestion), and *angashoonatha* (swelling) [19]. If we consider this as the criteria for diagnosing *amavata*, there are two major issues that a clinician may confront. The first issue is related to the operational definitions of some of the terms like *apaka*, which needs suitable questions pertaining to the construct, to be elicited in a given patient. Secondly, few of the *lakshanas* are present in other diseases as well, necessitating an algorithmic approach to differentiate this disease from overlapping ones. Hence it is imperative to formulate a standardized case definition or diagnostic criteria for each disease before the tool is being developed. Apart from diagnosing the disease, a diagnostic tool also should incorporate items aimed at categorizing it into various subtypes with regards to *dosha* status and disease staging. Such classification criteria can be incorporated along with the primary diagnosis or as a subsequent step in assessment. Framing the diagnostic and classification criteria may be accomplished through a multistep process including literature review, focus group discussions as well as consensus methods involving experts, experienced practitioners, and academicians. Consensus methods allow for a group of experts to share ideas to form a consensus on selected topics and include approaches like consensus conference, modified Delphi survey, and Nominal group technique [20,21].

7.2. The second phase – tool development and validation

This process is to be carried out in the following stages.

7.2.1. Devising items and response scales

Once the diagnostic criteria are set, the next step is to formulate questions for each of the elements or construct. Since many disease symptoms are expressed in intricate Sanskrit terminologies, an operational definition, as well as questions to elicit these in patients, are essential, to avoid discrepancies in their clinical application. Questions can be generated from several sources including literature review, expert interview, or by a focus group discussion involving experts as well as proposed respondents. In this process, terminologies from the classical textbooks may be taken up for discussion and necessary modifications done based on the panel inputs. The decision also has to be taken regarding the intended use of the tool, the number of items or questions needed to elicit each *lakshana*, and the type of response scales for each question. Necessary care should be taken regarding the relevance of each item, chronology, and wording of questions, and selection of response formats, to carry out the patient assessment in a stepwise manner. It is ideal to have a mixture of both positively and negatively worded items to minimize the danger of acquiescent response bias, i.e. the tendency for respondents to agree with a statement or respond in the same way to items [22].

7.2.2. Selection of response scales and formats

While framing the questions for health assessment, there is a wide variety of response scales that can be employed, depending upon the clinician's need as well as the factor being assessed. The most commonly used formats are based either on dichotomized categories (e.g. Yes or No format) or a continuum, which offers a range of choices to select from. Scales that allow patients to choose the option on a continuum agreement are preferred so that grading of a factor also can be assessed along with its presence. Such scales are commonly developed based on the Likert scale, Thurstone's method, or Guttman scaling [23]. For example, the Likert scale uses a bipolar scaling method ranging from agreement to disagreement or positive to negative statements, where items are rated 1–5 or 1–7. In the case of Ayurveda, this can be employed in assessing subjective parameters like *ruja* (pain), *kandu* (pruritus), or *daha* (burning sensation). [24]. Guttman scale utilizes questions or statements in a hierarchical order so that a respondent agreeing on a particular item will also agree with the lower order statements stated below that. As an example, questions on different stages of *dosha* vitiation can be arranged in a hierarchical order so that positive response for a given item will provide a cumulative score, indicating the staging in the given patient. Similarly, in the case of composite tools, some components may have a greater role in identifying a particular construct compared with others; where assigning weights to these components might become necessary [25]. Moreover, many of the scale items may have a level of inter-correlation as they aim to evaluate the same characteristic, which also significantly impacts its accuracy in predicting the outcome.

7.2.3. Pretesting

The prepared questionnaire has to be pretested in a small sample of 10–30 of the target population as well as in experts, to refine the questions. This step includes the assessment of face and content validity, translation review, reliability assessment, and item revision [26,27]. Another important step is to assess the respondent and interviewer friendliness of certain questions, or the entire questionnaire, which can be achieved through focus groups, cognitive interviews with test subjects, or both [28].

7.2.3.1. Face and content validity. After preparing the initial draft, it has to be examined by a few experts for its face validity and content validity. In the context of an instrument or tool, validity expresses the degree to which it measures the particular attribute or

construct, or it can be considered as “the appropriateness of inference or decision made from measurement”. Validity tests can be broadly classified into those that assess the theoretical construct and the empirical construct [29]. Face validity and content validity assess the theoretical construct whereas the empirical construct is assessed by means of criterion validity and construct validity, which is explained later on (Section 7.2.4).

Face validity: Face validity refers to the extent to which one or more experts subjectively agreeing that the items in the questionnaire are a valid measure of the concept which is being measured “just on the face of it”. It is often considered as very casual and soft so that many researchers do not consider this as a good indicator of validity. There are also others who consider that face validity is a component of content validity [30].

Content validity: Content validity indicates whether the scale items represent the proposed domains or concepts, the questionnaire is intended to measure [31]. This is a more reliable measure than face validity since it can be assessed by employing objective techniques. According to Bums and Grove, content validity can be “obtained from three sources: literature, representatives of the relevant populations, and experts”, which in turn could be established in two stages; development and judgment stage [32]. In the development stage, content validity is assessed through inputs from literature, population representatives, or experts, mainly employing qualitative methods like a survey or focus groups, as described earlier. In the judgment stage, content validity is assessed using an objective method, where graded responses are elicited from experts to generate quantitative evidence. Even though there are several methods through which this can be achieved, a simple technique is the Content validity index developed by Wal and Bausell [33]. Here the questionnaire is reviewed by five to ten experts, to judge the content domains of the tool, through the use of rating scales so that each item is ranked on a four-point scale, based on relevance, clarity, simplicity, and ambiguity see Table 2.

The content validity index is calculated from this data, which in turn will provide information about the level of agreement among experts regarding the items in the questionnaire. However, just like face validity, content validity also has a drawback, as it involves subjective assessment on the part of experts about the relevance of each item in the tool.

7.2.3.2. Cognitive interview. In self-reported questionnaires, it is important to assess whether the respondent interprets the items consistently with intended meanings [34]. This is carried out by cognitive interviews, where the questionnaire is administered to a selected number of subjects and their response elicited. Two important methods used are the Think aloud method and the Graded response scale, by which parameters like comprehension, retrieval, judgment, and response are assessed. This process helps in identifying confusing questions or response options and misinterpretation by the respondents, so that such questions can be refined or reframed to convey the intended meaning precisely.

7.2.3.3. Translation and back-translation. If the original instrument is developed in a different language, then it has to be translated into the one that will be used in practice and then back translate, to check if it keeps the original meaning after translation [35]. In such a scenario, the cognitive interview should be carried out with the translated version of the tool. The translation and back translation have to be carried out by two different experts respectively, to avoid bias.

7.2.3.4. Reliability assessment. Reliability is the extent to which a questionnaire, test, or any measurement procedure produces the same results on repeated attempts, provided the construct being

Table 2
Rating scale to calculate the content validity index.

Criteria	Rank			
	1	2	3	4
1. Relevance	Not relevant	The item needs some revision	Relevant but need minor revision	Very relevant
2. Clarity	Not clear	The item needs some revision	Clear but need minor revision	Very clear
3. Simplicity	Not simple	The item needs some revision	Simple but need minor revision	very simple
4. Ambiguity	Doubtful	The item needs some revision	No doubt but need minor revision	Meaning is clear

assessed is stable over time [36]. It is a very important property of any diagnostic tool, especially in a clinical trial, because it is vital to establish that any changes observed in the trial are due to the intervention and not to errors in the measuring instrument. Reliability is not a fixed property of a questionnaire as it depends on the intended function, the population on which it is applied, and the conditions in which it is used, so that the same instrument may not be reliable under different conditions. There are three aspects of reliability, namely: stability (test-retest reliability), equivalence (inter-observer reliability), and internal consistency [30]. These are discussed below:

a. Stability - test-retest reliability: Test-retest reliability indicates the stability of the measurement tool over time so that similar scores are obtained when the test is repeated on the same subjects at a different timepoint, provided the construct is stable. The extent to which these repeated values are similar to one another reflects the test-retest reliability of that measure [37].

Intra-rater-reliability: A variant of the test-retest reliability is the intra-rater reliability, where the same observer makes two separate measurements in the same subjects, at separate moments in time and a comparison is made [38]. This is especially done if the assessment involves a human component in decision making and is a measure of the stability of scores given by the same evaluator on repeated attempts. Statistical assessment involves estimation of the correlation between multiple measurements, with the use of intraclass correlation coefficient for continuous measures, Spearman rank correlation coefficient for ordinal measures, and phi coefficient in the case of binary variables. In general, the correlation coefficient (r) values of $r \geq 0.70$ are considered as good, indicating the stability of the instrument [39]. However, this form of reliability cannot be assessed in certain cases where the factor that is measured keeps changing with time. Hence two important assumptions have to be met to use this test-retest procedure. The first and most important assumption is that the characteristic measured does not change over time, called a testing effect. The second one is that the interval between assessments is long enough yet short in time so that the respondent's memory of the previous test does not influence the second attempt (memory effect). Streiner and Norman (2015) suggest that the usual range of time elapsed between assessments tends to be between 2 and 14 days [39].

b. Equivalence: Equivalence refers to the consistency of the results among multiple administrators of an instrument, or among alternate forms of the same instrument. The former is called inter-observer or inter-rater reliability whereas the latter is known as alternate-form reliability [40].

Inter-rater reliability: This form of reliability is assessed by comparing measurements obtained by two or more observers at a given time, on the same population and the agreement between them denotes inter-rater reliability. This is important in the case of ayurvedic diagnosis, especially when the tool includes the assessment of constructs that necessitates a physician's judgement, as in the case of categorizing different shades of colour of the affected body part in diseases like *sopha* (oedema) or *vrana* (ulcer). However, it is to be noted that Intra-rater reliability is a prerequisite for

inter-rater reliability. Reliability studies that measure agreement between two or more observers usually make use of Kappa statistic (Cohen's or Fleiss kappa) where a kappa score of 1 indicates perfect agreement, while zero indicates agreement equivalent to chance [41]. For conducting such reliability studies, there has to be a sufficient sample size and also requires blinding of observers, in order to prevent bias in assessment. In alternate-forms reliability, different versions (with altered wordings) of the same tool are administered and evaluated for the degree of correlation between the assessments [47]. However, in practice, it is rarely employed owing to the difficulty in framing and administering multiple questionnaires.

c. Internal consistency (homogeneity): Normally, in a questionnaire, more than one item or question is used to measure an attribute or construct, because of the basic tenet that several related observations can produce a more reliable estimate than one. For example, a *prakriti* assessment tool may require more than one question to measure a single attribute "*agni*", among others. Internal consistency measures the extent to which items in the test or instrument are measuring the same attribute of the given construct [42]. Several methods can be employed for measuring internal consistency that includes item-to-total correlation, split-half reliability, Kuder–Richardson coefficient, and Cronbach's α . The item–total correlation is used in assessing the reliability of a multi-item scale where the correlation between an individual item and the total score without that item is calculated, so that odd questions are singled out [43]. In split-half reliability, the results are divided into half so that correlations are calculated comparing both halves and strong correlations indicate high reliability [44]. Another commonly used measure is an extension of split-half reliability, termed Coefficient alpha or Cronbach's alpha, which estimates the average level of agreement in all the possible ways of split-half tests and higher alpha indicating good internal consistency [45]. This is used in the case of scales with items that have several response options, whereas, in tools with dichotomous response scales, the more complicated Kuder–Richardson test is employed [46]. Currently, all the above measures can be carried out using available statistical software, so that the selection of the test is the only factor that requires prudence on the part of the researcher.

7.2.3.5. Item revision. During the process of pretesting, if the instrument is found to have a poor face and content validity or low reliability or if there is a poor understanding of the items and responses scales by the target population, then the respective items have to be revised or deleted, to refine the tool. Following the item revision, the questionnaire has to be retested for the above parameters, before administering to a larger population for assessing its validity and other accuracy measures.

7.2.4. Empirical validation - large sample study

After revising the items, the validity of the questionnaire can be empirically established with the help of a field test, to assess how well the given measure relates to one or more external criterion or the intended constructs. These forms of validity are called criterion-

related validity and construct validity respectively [30]. Further, the criterion validity is divided into predictive validity and concurrence validity whereas the subtypes of construct validity include convergence validity, discriminant validity, known-group validity, and factorial validity. Some experts have also included hypothesis-testing validity as a form of construct validity [47]. In this phase, the tool has to be administered in a large sample, with sample size being calculated based on the number of items in the questionnaire.

7.2.4.1. Criterion validity. Establishing the criterion validity involves the demonstration of a correlation between the new tool and another instrument or standard that is considered as an accurate indicator of the same concept or construct being measured (the gold standard) [48]. A major disadvantage of this validity is that such a gold standard may not be available or easy to establish, as in the case of Ayurveda. Criterion-related validity is further classified into concurrent validity and predictive validity [49]. Concurrent validity is assessed statistically by testing the new instrument against an independent criterion or existing standard, where both tools are administered on the same subjects, for calculating the correlation coefficient and further tested using Two-sided t-tests (TOST) [50]. In case if such a criterion or standard is absent, as in most of the Ayurvedic diagnosis, this correlation can be done with a panel diagnosis involving experts. On the other hand, predictive validity is assessed when the purpose of the tool is to predict or estimate the occurrence of a behaviour or an event and is often described in terms of sensitivity and specificity [51], as explained later in the third phase.

7.2.4.2. Construct validity. Construct validity is a quantitative form of assessing the degree to which the tool measures the trait or theoretical construct that it is intended to measure [52]. For example, in the case of health assessment tools, most often the instrument is intended to measure an underlying construct, such as pain, or disability, rather than some directly observable phenomenon. Such constructs that are not directly measurable can be expected to have a set of quantitative relationships with other constructs (e.g. exercise tolerance) as per current understanding. Since a single related observation cannot prove the construct validity of a new measure, multiple variables may be used for this purpose such as disease staging, clinical or laboratory evidence, or other related constructs of well-being. As this assessment uses a hypothetical construct for comparison, it is the most difficult one to establish, despite being the more valuable form of validity. Depending upon the purpose of the tool, there are several means of evidence that can be used for establishing the construct validity, as discussed below:

Convergent and discriminant validity: These are two sophisticated forms of testing construct validity, which requires postulating that the instrument under consideration should have stronger relationships with some variables and weaker relationships with others [53]. Accordingly, correlations are expected to be strongest with the most related constructs and weakest with the most distally related constructs. Thus, while assessing the given construct with the new tool, the result is compared with a different measure of the same concept, so that if both yield similar results, the validity is established. Similarly, discriminant validity verifies that the given instrument measures a construct that is not related to the one under consideration. For example, a new tool assessing *swasthya* (positive health) will have convergent validity with tools measuring general health like WHOQOL-BREF (World Health Organization Quality of Life Instruments), whereas discriminant validity with that assessing disability like WHODAS 2.0. (World Health Organization Disability Assessment Schedule).

7.2.4.3. Factorial validity/factor analysis. Even though factorial validity is an empirical extension of content validity, it is considered as a subtype of construct validity, because it employs a statistical model called factor analysis to validate the contents of the construct [54]. This form of validity is assessed in cases where the construct of interest has several dimensions or if the instrument has different domains of a general attribute. For example, the tool measuring *swasthya* will be multi-dimensional, so that it needs to assess the physical, psychological and social aspects of an individual. In such a case, items set up for measuring a particular domain (e.g. physical) within the construct of interest (overall health status), are supposed to be highly related to one another, than to those measuring other dimensions (psychological or social domains). In the process of factor analysis, the items are analysed creating a mathematical model that estimates construct domains, within the pool of items. It assesses the intercorrelation between questions or up to what degree individual items are measuring a common factor or domain so that items with poor factor loading can be deleted from the tool. The main statistical methods used here are correlation coefficients like Pearson's or Spearman's and principal component analysis [55,56].

7.3. Third phase - diagnostic test assessment

Before a new clinical tool can be introduced into general practice, it should be evaluated for its clinical validity, by the means of diagnostic accuracy studies. These studies evaluate the new test's accuracy in discriminating subjects with or without the target condition, in comparison with an existing standard, and can be considered as a prototype of criterion validity explained above. Several parameters are assessed in this step including sensitivity, specificity, predictive values, likelihood ratios, and area under the ROC curve [57]. Various study designs can be employed for this purpose, based on the condition and population of interest, including Cohort (single gate entry), Case-control (two gate entry), and randomized controlled trials [58,59]. The parameters to be assessed in this phase are briefly described below.

7.3.1. Sensitivity and specificity

If the instrument is developed to detect the presence or absence of a particular phenomenon or disease, it is important to determine the degree of agreement between the results obtained by the new tool (index test) and an existing gold standard. In case if the new scale is a continuous measure and the external criterion a dichotomous one (presence or absence of disease), then it is imperative to choose a cut point that can classify the subjects as healthy or sick. This can lead to two types of errors: healthy individuals labelled as sick (false positive) or sick subjects diagnosed as healthy (false negative). Assessing the sensitivity and specificity will help in quantifying the diagnostic ability of the tool, especially in differentiating subjects with and without the particular disease [60]. The sensitivity of a diagnostic tool is defined as the proportion of people with the target condition who got a positive test result. In other words, this indicates the ability of the tool to detect subjects with the disease, whereas, specificity is the proportion of people without the target condition, among those who tested negative. This indicates the ability of the tool to rule out the disease in most of the healthy subjects. In general, a diagnostic test is considered to have a reasonable validity if its sensitivity and specificity are equal or over 0.80 [61].

7.3.2. Predictive values

In addition to the assessment of diagnostic validity, it is also important to evaluate its behaviour, when applied to different

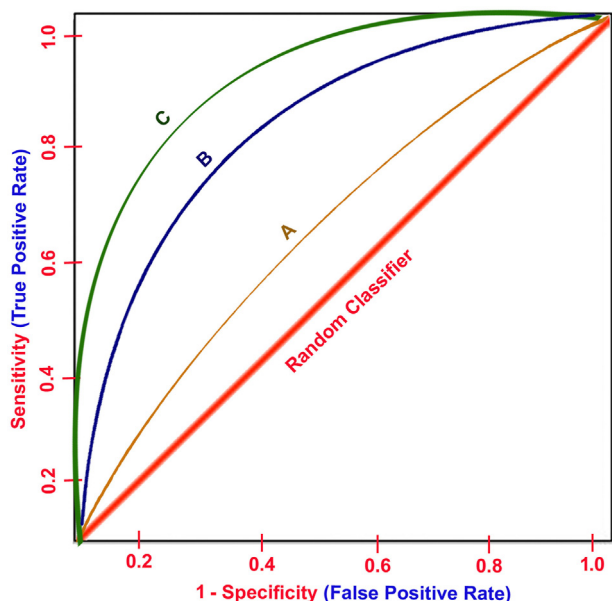


Fig. 2. ROC curve for three different cut points denoted by A, B, and C. Compared to A and B, C represents the best classifier among all the three cut-offs.

clinical contexts. This is done by calculating the predictive values, which also takes into consideration the population to which it is applied as well as the prevalence of the disease in them [60]. Predictive values are classified into two, where the positive predictive value (PPV) indicates the proportion of patients who test positive and actually have the disease, whereas negative predictive value (NPV) is the proportion of patients who test negative and are truly free of the disease. It is to be noted that, unlike sensitivity and specificity, the PPV and NPV are dependent on the population being tested and the prevalence of the disease. If a particular disease is very common in the given population, then the calculated PPV will be high, indicating that if the test is positive, the patient is certain to have that disease.

7.3.3. Likelihood ratio

It has been already stated that predictive values take prevalence into consideration, which can influence the validity of a diagnostic tool. One way to avoid this influence is to calculate likelihood ratios, that relate sensitivity and specificity in a single index, without considering the prevalence of the disease.

The likelihood ratio is defined as how much more likely a patient who tests positive has the disease compared with one who tests negative and demonstrates the potential utility of the diagnostic tool [59].

Table 3 Summarizing the phases of tool development.

Phases in Tool Development	Steps	Methods
Preliminary Phase (Defining Diagnostic Criteria)	Defining the diagnostic / classification criteria	1. Literature Review 2. Focus Group Discussions 3. Consensus Methods: a. Consensus Conference b. Modified Delphi c. Nominal Group Technique
Second Phase (Tool Development and Validation)	Devising Items & Response Scales Item generation 1. Number of Items 2. Type of response 3. Selection of response scales and formats	1. Literature review 2. Focus group discussions 3. Modified Delphi 4. Dichotomous 5. Continuous a. Thurstone's method b. Likert scale c. Guttman scale
	Pretesting Face Validity Content Validity Cognitive Interview Translation & Back translation Reliability assessment	Expert evaluation Small sample study in respondents Language experts Small sample study Internal consistency (Cronbach's α) Test-Retest Reliability Inter-Rater Reliability (Kappa statistics)
	Item revision Empirical Validation and reliability assessment- Criterion Validity Construct Validity Factor Analysis	Large sample study Correlating with gold standard or adopt measures for missing gold standard Convergent Validity Discriminant Validity Pearson's or Spearman's coefficients Principal component analysis Cohort (single gate entry) Case-control (two gate entry) Randomized controlled trials
Third Phase (Diagnostic Test Assessment)	Sensitivity Specificity Predictive Values Likelihood Ratios Area under ROC curve Construct reference standard	Composite reference standard Panel Diagnosis Latent class analysis <i>Upasaya&Anupasaya</i> Other outcome measurements
*Missing Gold Standard	Validate Index test results	

The likelihood ratio for a positive result (LR+): It is calculated by dividing the proportion of sick subjects with a positive test result (sensitivity) by the proportion of healthy subjects with a negative result (1-specificity).

The likelihood ratio for a negative result (LR-): It is the probability of a person with the disease tested negative (false negative) divided by the probability of a person without the disease and tested negative (true negative). Of these 2 indices, the likelihood ratio for a positive result is the most commonly employed in practice, known as the “likelihood ratio.”

7.3.4. ROC curve

When the values of a diagnostic test follow a quantitative scale, sensitivity and specificity vary according to the chosen cut point that classifies the population as healthy or sick. In this situation, a global measure of the validity of the test in the universe of all possible cut points is obtained through the use of ROC curves (receiver operating characteristics) (Fig. 2) [62].

The ROC curve is drawn by plotting sensitivity along the Y-axis and the complement of specificity (1-specificity) along the X-axis. Different curves are then plotted for each cut-off point and the cut off with the maximum area under the curve (AUC) correctly classifies the pair of sensitivity and specificity ideal for the developed tool see Table 3.

8. The problem of missing gold standard

Establishing the reliability and validity measures of a tool requires evaluation against an existing gold standard, which may not always be feasible in Ayurveda. In such a scenario, several methods could be adopted, depending upon the characteristics of the existing standards. These include imputing or adjusting for missing data, generating a construct reference standard, and validating the index test results in relation to other relevant clinical characteristics [63]. Among the above-mentioned methods, the last two can be adopted in Ayurveda. For instance, results from multiple methods like composite reference standard [64], panel diagnosis [65], and latent class analysis [66] can be combined to generate a construct reference standard. In Ayurveda, a panel diagnosis in the form of expert consensus is more ideal since the diagnosis involves a greater human element. Another alternative method is to abandon the diagnostic accuracy paradigm and validate the index test results in relation to other characteristics like future clinical events or outcomes. Whether the approach of *upasaya* and *anupasaya* (diagnosis based on explorative therapy) [67] expounded in Ayurveda can be employed for validating the index test results is also to be explored. However, unlike the diagnostic accuracy indicators, analysis of these studies incorporates the use of measures like event rates, relative risks, and other correlation statistics.

9. Other recommendations

Developing diagnostic instruments in compliance with the contemporary methods will aid in standardizing the ayurvedic diagnostic approach, without relying on the biomedical methods of assessment. The whole process also solicits extensive inputs from allied fields like statistics and social sciences, aiding the investigator to effectively integrate the principles of psychometry in tool development. However, not all measures of validity and reliability explained here are essential in every instance, as the selection of specific psychometric property depends upon the type and purpose of the tool, as well as the population in which it is being administered. Likewise, these measures are not constant for a given tool and require separate assessments for different settings. Moreover, there may be settings where a modern diagnosis is expedient in

patient management, especially in determining the prognosis or evaluating the outcome. Hence efforts could also be directed at formulating a framework for integrating the modern diagnostic measures or investigations within the ambit of ayurvedic diagnosis. Apart from this, it is also worthwhile to examine whether the contemporary concepts of validity and reliability can be compared with the classical research paradigms elucidated in Ayurveda such as *Pramanas* [68]. Such an attempt will address the long-term demand for developing tools and assessment methods within the context of ayurvedic theoretical constructs.

10. Conclusion

Ayurveda, with its holistic and person centric clinical approach, relies on the assessment of subjective and objective parameters for arriving at an individualized diagnosis. Further, when attempting a personalized diagnosis in a given patient, it is imperative to have a certain degree of agreement between clinicians, which can be brought about by employing standardised diagnostic tools. Currently, there are no widely accepted standards for diagnostic tool development in Ayurveda. The authors, after reviewing the current literature, propose a framework for tool development in Ayurveda, which involves three phases viz. defining the diagnostic criteria, tool development and validation, and diagnostic test assessment. The methodological challenges like the interplay of multiple variables in the diagnosis and lack of a gold standard for comparison were also discussed with their probable solutions.

Source(s) of funding

None.

Conflict of interest

None.

Acknowledgements

The authors would like to extend their gratitude towards Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya, Kanchipuram, Tamil Nadu, for providing resources and support.

References

- [1] Thakar VJ. Diagnostic methods in ayurveda. *Ancient Sci Life* 1982 Jan;1(3):139.
- [2] Kurande VH, Waagepetersen R, Toft E, Prasad R. Reliability studies of diagnostic methods in Indian traditional Ayurveda medicine: an overview. *J Ayurveda Integr Med* 2013 Apr;4(2):67.
- [3] Niemi M, Stähle G. The use of ayurvedic medicine in the context of health promotion—a mixed methods case study of an ayurvedic centre in Sweden. *BMC Compl Alternative Med* 2016 Dec 1;16(1):62.
- [4] Manohar PR. Consideration of Ayurvedic diagnostics in design of clinical trials. *Ancient Sci Life* 2013 Jul;33(1):1.
- [5] Manohar PR. Clinical diagnosis in ayurveda: challenges and solutions. *Ancient Sci Life* 2012 Apr;31(4):149.
- [6] Jutel A. Sociology of diagnosis: a preliminary review. *Social Health Illness* 2009 Mar;31(2):278–99.
- [7] McNeil BJ, Keeler E, Adelstein SJ. Primer on certain elements of medical decision making. *N Engl J Med* 1975 Jul 31;293(5):211–5.
- [8] Coste J, Fermanian J, Venot A. Methodological and statistical problems in the construction of composite measurement scales: a survey of six medical and epidemiological journals. *Stat Med* 1995 Feb 28;14(4):331–45.
- [9] Wang CH, Hsueh IP, Sheu CF, Hsieh CL. Discriminative, predictive, and evaluative properties of a trunk control measure in patients with stroke. *Phys Ther* 2005 Sep 1;85(9):887–94.
- [10] Bolboacă SD. Medical diagnostic tests: a review of test anatomy, phases, and statistical treatment of data. *Computat Math Method Med* 2019 May 28:2019.
- [11] Lukas A, Niederecker T, Günther I, Mayer B, Nikolaus T. Self-and proxy report for the assessment of pain in patients with and without cognitive impairment. *Z Gerontol Geriatr* 2013 Apr 1;46(3):214–21.

- [12] Vaidya Jadavaji TA, editor. Charaka Samhita of Acharya Charaka, Vimana Sthana, Rogabhishagjithiyam, 8th Chapter, Verse 94. 5th ed. Varanasi: Chaukambha Krishnadas Academy; 2001. p. 276.
- [13] Vaidya Jadavaji TA, editor. Susrutha samhitha of Acharya susrutha, Sutrasthana, Vranaprasnam adhyaya, 22nd Chapter, Verse 36. Varanasi: Chaukambha Orientalia; 2014. p. 106.
- [14] Upadhyaya Yadunandana, editor. Madhavanidana of Madhavkara, Panchanidanalakshana, 1st chapter, verse 5-6. Varanasi: Chowkambha Sanskrit Sanstan; 1993. p. 36.
- [15] Souza AC, Alexandre NM, Guirardello ED. Psychometric properties in instruments evaluation of reliability and validity. *Epidemiologia e Serviços de Saúde* 2017;26:649–59.
- [16] Avlund K, Schultz-Larsen K, Kreiner S. The measurement of instrumental ADL: content validity and construct validity. *Aging Clin Exp Res* 1993 Oct 1;5(5): 371–83.
- [17] Baerheim A. The diagnostic process in general practice: has it a two-phase structure? *Fam Pract* 2001;18(3):243–5.
- [18] Aggarwal R, Ringold S, Khanna D, Neogi T, Johnson SR, Miller A, et al. Distinctions between diagnostic and classification criteria? *Arthritis Care Res* 2015 Jul;67(7):891.
- [19] Upadhyaya Yadunandana, editor. Madhavanidana of Madhavkara, Amavatanidana, 25th chapter, verse 5-6. Varanasi: Chowkambha Sanskrit Sanstan; 1993. p. 508.
- [20] Bourrée F, Michel P, Salmi LR. Consensus methods: review of original methods and their main alternatives used in public health. *Revue d'epidemiologie et de sante publique* 2008 Dec 1;56(6):e13–21.
- [21] Kea B, Sun BC. Consensus development for healthcare professionals. *Intern Emerg Med* 2015 Apr 1;10(3):373–83.
- [22] Schriesheim CA, Hill KD. Controlling acquiescence response bias by item reversals: the effect on questionnaire validity. *Educ Psychol Meas* 1981 Dec;41(4):1101–14.
- [23] Streiner DL, Norman GR, Cairney J. Health measurement scales: a practical guide to their development and use. USA: Oxford University Press; 2015. p. 38–67.
- [24] Voutilainen A, Pitkääho T, Kvist T, Vehviläinen-Julkunen K. How to ask about patient satisfaction? The visual analogue scale is less vulnerable to confounding factors and ceiling effect than a symmetric Likert scale. *J Adv Nurs* 2016 Apr;72(4):946–57.
- [25] Panagiotakos D. Health measurement scales: methodological issues. *Open Cardiovasc Med J* 2009;3:160.
- [26] Presser S, Couper MP, Lessler JT, Martin E, Martin J, Rothgeb JM, et al. Methods for testing and evaluating survey questions. *Publ Opin Q* 2004 Mar 1;68(1):109–30.
- [27] Howard MC. Scale pretesting. *Practical Assess Res Eval* 2018;23(1):5.
- [28] Giesen D, Meertens V, Vis-Visschers R, Beukenhorst D. Questionnaire development. The Hague, Heerlen, Netherlands. 2012. p. 36–45.
- [29] Zamanzadeh V, Rassouli M, Abbaszadeh A, Majd HA, Nikanfar A, Ghahramanian A. Details of content validity and objectifying it in instrument development. *Nurs Prac Today* 2014;1(3):163–71.
- [30] Bolarinwa OA. Principles and methods of validity and reliability testing of questionnaires used in social and health science researches. *Niger Postgrad Med J* 2015 Oct 1;22(4):195.
- [31] Haynes SN, Richard D, Kubany ES. Content validity in psychological assessment: a functional approach to concepts and methods. *Psychol Assess* 1995 Sep;7(3):238.
- [32] Roberts P, Priest H. Reliability and validity in research. *Nurs Stand* 2006 Jul 12;20(44):41–6.
- [33] Yaghmale F. Content validity and its estimation. *J Med Educ* 2003;3(1):25–7.
- [34] Castillo-Díaz M, Padilla JL. How cognitive interviewing can provide validity evidence of the response processes to scale items. *Soc Indic Res* 2013 Dec 1;114(3):963–75.
- [35] Del Greco L, Walop W, Eastridge L. Questionnaire development: 3. Translation. *CMAJ (Can Med Assoc J)* 1987 Apr 15;136(8):817.
- [36] Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. *Health Technol Assess* 1998;2(14).
- [37] Hendrickson AR, Massey PD, Cronan TP. On the test-retest reliability of perceived usefulness and perceived ease of use scales. *MIS Q* 1993 Jun 1:227–30.
- [38] Nutter Jr FW, Gleason ML, Jenco JH, Christians NC. Assessing the accuracy, intra-rater repeatability, and inter-rater reliability of disease assessment systems. *Phytopathology* 1993 Aug 1;83(8):806–12.
- [39] Streiner DL, Norman GR, Cairney J. Health measurement scales: a practical guide to their development and use. USA: Oxford University Press; 2015. p. 172–3.
- [40] LoBiondo-Wood G, Haber J. Reliability and validity. *Nursing research. Methods and critical appraisal for evidence-based practice*. 2014 Mar 12. p. 289–309.
- [41] McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med* 2012 Oct 15;22(3):276–82.
- [42] Rattray J, Jones MC. Essential elements of questionnaire design and development. *J Clin Nurs* 2007 Feb;16(2):234–43.
- [43] Hwang IH. The usability of item-total correlation as the index of item discrimination. *Korean J Med Educ* 2000 Jun 1;12(1):45–51.
- [44] Streiner DL. Starting at the beginning: an introduction to coefficient alpha and internal consistency. *J Pers Assess* 2003 Feb 1;80(1):99–103.
- [45] Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int J Med Educ* 2011;2:53.
- [46] Cliff N. An improved internal consistency reliability estimate. *J Educ Stat* 1984 Jun;9(2):151–61.
- [47] O'Leary-Kelly SW, Vokurka RJ. The empirical assessment of construct validity. *J Oper Manag* 1998 Jul 1;16(4):387–405.
- [48] Rauta S, Salanterä S, Vahlberg T, Junttila K. The criterion validity, reliability, and feasibility of an instrument for assessing the nursing intensity in peri-operative settings. *Nurs Res Pract* 2017 Jul 17:2017.
- [49] Drost EA. Validity and reliability in social science research. *Educ Res Perspect* 2011;38:105–23.
- [50] Lakens D, Scheel AM, Isager PM. Equivalence testing for psychological research: a tutorial. *Adv Method Pract Psychol Sci* 2018 Jun;1(2):259–69.
- [51] Morisky DE, Green LW, Levine DM. Concurrent and predictive validity of a self-reported measure of medication adherence. *Med Care* 1986 Jan 1:67–74.
- [52] Colliver JA, Conlee MJ, Verhulst SJ. From test validity to construct validity... and back? *Med Educ* 2012 Apr;46(4):366–71.
- [53] Corral-Verdugo V, Figueroa AJ. Convergent and divergent validity of three measures of conservation behavior: the multitrait-multimethod approach. *Environ Behav* 1999 Nov;31(6):805–20.
- [54] Asmundson GJ, Bovell CV, Carleton RN, McWilliams IA. The fear of pain questionnaire—short form (FPQ-SF): factorial validity and psychometric properties. *Pain* 2008 Jan 1;134(1–2):51–8.
- [55] Williams B, Onsman A, Brown T. Exploratory factor analysis: a five-step guide for novices. *Australas J Paramed* 2010;8(3).
- [56] Swisher LL, Beckstead JW, Bebeau MJ. Factor analysis as a tool for survey analysis using a professional role orientation inventory as an example. *Phys Ther* 2004;84(9):784–99.
- [57] Gleason PM, Harris J, Sheean PM, Boushey CJ, Bruemmer B. Publishing nutrition research: validity, reliability, and diagnostic test assessment in nutrition-related research. *J Am Diet Assoc* 2010;110(3):409–19.
- [58] Knottnerus JA, Muris JW. Assessment of the accuracy of diagnostic tests: the cross-sectional study. *J Clin Epidemiol* 2003;56(11):1118–28.
- [59] Weinstein S, Obuchowski NA, Lieber ML. Clinical evaluation of diagnostic tests. *Am J Roentgenol* 2005;184(1):14–9.
- [60] Akobeng AK. Understanding diagnostic tests 1: sensitivity, specificity and predictive values. *Acta Paediatr* 2007;96(3):338–41.
- [61] Lalkhen AG, McCluskey A. Clinical tests: sensitivity and specificity. *Cont Educ Anaesth Crit Care Pain* 2008;8(6):221–3.
- [62] Mossman D, Somoza E. ROC curves, test accuracy, and the description of diagnostic tests. *J Neuropsychiatry Clin Neurosci* 2007;96(3):338–41.
- [63] Reitsma JB, Rutjes AW, Khan KS, Coomarasamy A, Bossuyt PM. A review of solutions for diagnostic accuracy studies with an imperfect or missing reference standard. *J Clin Epidemiol* 2009;62(8):797–806.
- [64] Naaktgeboren CA, Bertens LC, van Smeden M, de Groot JA, Moons KG, Reitsma JB. Value of composite reference standards in diagnostic research. *BMJ* 2013;347:f5605.
- [65] Bertens LC, Broekhuizen BD, Naaktgeboren CA, Rutten FH, Hoes AW, van Mourik Y, et al. Use of expert panels to define the reference standard in diagnostic research: a systematic review of published methods and reporting. *PLoS Med* 2013;10(10):e1001531.
- [66] van Smeden M, Naaktgeboren CA, Reitsma JB, Moons KG, de Groot JA. Latent class models in diagnostic studies when there is no reference standard—a systematic review. *Am J Epidemiol* 2014;179(4):423–31.
- [67] Vaidya Jadavaji TA, editor. Charaka Samhita of Acharya Charaka, Nidana Sthana, Jvara Nidana, 1st Chapter, Verse 10. 5th ed. Varanasi: Chaukambha Krishnadas Academy; 2001. p. 195.
- [68] Vinodkumar MV, Anoop AK. Review on comparability of 'classical' and 'contemporary' research methods in the context of Ayurveda. *J Ayurveda Integr Med* 2019.