Contents lists available at ScienceDirect

Journal of Ayurveda and Integrative Medicine

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

Original Research Article (Clinical)

Reliability of self-reported constitutional questionnaires in Ayurveda diagnosis

CrossMark

Corina Dunlap [a], [*], Douglas Hanes [a], Charles Elder [b], Carolyn Nygaard [a], Heather Zwickey [a]

[a] Helgott Research Institute c/o National University of Natural Medicine, 2220 SW 1st Ave, Portland, OR 97201, USA
[b] Kaiser Permanente Center for Health Research, 3800 N. Interstate Ave, Portland, OR 97227, USA

ABSTRACT

Background: Ayurveda is one of the most ancient and widely practiced forms of medicine today, along with Traditional Chinese Medicine. It consists of determining an individual's constitution or Prakriti and current imbalance(s) through the use of multimodal approaches. Ayurveda practitioners may choose to include either a self-reported or structured interview constitutional questionnaire as part of the Prakriti assessment. Currently, there is no standardized or validated self-reported constitutional questionnaire tool employed by Ayurveda physicians or western Ayurveda educational institutions.
Objectives: To examine test-retest reliability of three self-administered constitutional questionnaires at a one month interval and internal consistency of items pertaining to a single constitution.
Materials and methods: Three constitutional questionnaires were administered online. 19 participants completed three questionnaires at two time points, one month apart. Age range was 21–62 years old with a mean age of 34. Of the 19, 5 were male and 14 female. Vata, Pitta, and Kapha scores obtained from each questionnaire were standardized to give a vector of three relative percentages, summing to 100. These percentages were further translated from numerical values to one of ten possible dosha diagnoses.
Results: Analysis indicated that the three questionnaires had moderately good test-retest reliability according to numerical scores, but highly variable reliability according to discrete Ayurveda diagnosis. Internal consistency pertaining to individual constitutions within one questionnaire was poor for all three primary doshas, but especially for Kapha.
Conclusion: Further research is necessary to develop a reliable and standardized constitutional questionnaire.
© 2017 Transdisciplinary University, Bangalore and World Ayurveda Foundation. Publishing Services by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

The traditional Indian medical model known as Ayurveda is one of the most ancient and widely practiced forms of medicine today, along with Traditional Chinese Medicine. As demand in the western world for traditional medicine increases, there is a growing interest to ensure quality in training, research, and practice [1]. Treatment efficacy is the most prolific type of research in Ayurveda [1]; however there is little research examining reliability of the various diagnostic techniques upon which treatment prescription and efficacy depend. The Ayurveda diagnosis consists of determining an individual's constitution and current imbalance(s) through the use of a multimodal approach including observation, physical exam, pulse diagnosis, and health history. Many Ayurveda physicians and western Ayurveda schools also employ some version of a constitutional questionnaire during the initial patient intake. These questionnaires are often made available online and are popular tools for self-diagnosis amongst the general public. They are not however, standardized or have evidence of validity. If included in the overall assessment, these questionnaires may impact diagnosis and long-term treatment recommendations. Research on their reliability and validity is therefore imperative.

According to the philosophy of Ayurveda, humans have physical and behavioral differences that are classified into one or more of three metabolic forces, or doshas [2]. These doshas, known as vata, pitta, and kapha, are the vital bioenergies responsible for promoting and sustaining the health of each individual. Each dosha comprises

of five elements: earth, air, fire, water, and space. *Vata* is the combination of air and space, *pitta-* of fire and water, and *kapha-* of earth and fire.

An individual's specific *Prakriti* or constitution refers to the physical and behavioral qualities that remain stable throughout one's life [3]. Ayurveda considers seven *Prakriti* classifications, however, there are ten possible combinations depending on relative predominance of *dosha*: *vata, pitta, kapha, vata-pitta, pitta-vata, vata-kapha, kapha-vata, pitta-kapha, kapha-pitta,* or *Tridosha*; the *dosha* listed first being the more dominant of the two for an individual who is *Dvidoshic*. Each *Prakriti* classification describes the predominant *dosha*(s) that is likely to overpower the others, producing a certain set of characteristic physiologic imbalances. From an Ayurvedic point of view, knowing one's *Prakriti* enables a person to make educated lifestyle choices in order to minimize the effects of such inherent tendencies [4].

Vikriti on the other hand, is a term used to describe the changed condition of body, mind, and consciousness [4]. While *Prakriti* remains stable throughout one's life, *Vikriti* is a temporary state of imbalance of the *doshas*, constantly changing depending on one's lifestyle habits. It is imperative for the success of an Ayurveda treatment plan that the physician correctly diagnoses an individual's *Vikriti*; but, it is very helpful for the individual and the physician to know the underlying *Prakriti* or constitution as well as it can inform potential future imbalances, disease susceptibility, and long-term treatment plans [5,6]. When evaluating effectiveness of an Ayurveda prescription, for example, it is important to consider *Prakriti*. An appropriate constitutionally based prescription can enhance therapeutic effects and minimize adverse effects.

Traditionally, there are four methods of determining or diagnosing an individual's *Prakriti*: observation, physical exam, pulse diagnosis, and health history [4]. These methods used together are preferable to any used alone as individually, they may lead to bias in *Prakriti* diagnosis [7]. Several studies have examined reliability and validity of *Prakriti* diagnosis by way of various methods: pulse taking alone [1,6], interview-based or self-reporting questionnaires in combination and alone [8–10], and incorporating several of the diagnostic methods together [1,8,9]. To date, results of these studies have varied greatly, from low to moderate levels of reliability, and there is not a validated standard by which to compare one method to another.

Ayurveda practitioners may choose to include either a self-reported or structured interview constitutional questionnaire as part of the *Prakriti* assessment. At present, there is no standardized or validated self-reported constitutional questionnaire tool employed by Ayurveda physicians or western Ayurveda educational institutions. Rather, there are a wide variety of questionnaires, many of which are publicly available. It is our intention to contribute to the existing literature on examination of the reliability of self-reported constitutional questionnaires by investigating three publicly available questionnaires developed by two of the most well known western Ayurveda educational institutions and by a private international Ayurveda products company. We are unaware of any prior research investigating the reliability or validity of these specific self-reporting constitutional questionnaires. We chose these questionnaires as our starting point because these sources are easily available to the general public and influence western trained Ayurveda practitioners, leading many to employ these or very similar questionnaires in private medical practice. For purpose of this research, we have decided to focus solely on the subject of reliability of self-reported questionnaires without incorporating any other diagnostic methods (i.e. pulse, interview-style questionnaires) due to not having a validated standard by which to compare results.

Reliability refers to consistency and repeatability of outcome measures [1,11]. Test–retest reliability is used to assess consistency of measures between two points set apart by a length of time. It is necessary to separate the two measures by an adequate amount of time so that results are not influenced by the observer's memory. Based on methods used to assess test–retest reliability for various other self-reporting diagnostic questionnaires [10,14–16], we chose a 1-month time period to ensure an adequate length of time before retesting.

Internal consistency reliability refers to consistency of responses across individual items within a test that intend to measure the same construct [1]. Here, it is a way of assessing the amount of agreement amongst questions that examine the *vata, pitta,* and *kapha* diagnoses. If, for example, a survey uses a number of questions to assess *vata dosha* — i.e., a high score on these questions is supposed to identify presence of *vata dosha* in the respondent — then we would expect there to be high levels of agreement between answers to these questions.

It is our hope that, by exploring the structure and reliability of various self-reporting questionnaires used by western Ayurveda institutions, we can contribute to development of a reliable and validated self-reporting constitutional questionnaire tool that can be shared widely by Ayurveda educational institutions and practitioners.

2. Materials and methods

Participants were recruited by two methods: 1) an email sent to the local naturopathic school's student body and 2) flyers posted on the school's campus and throughout the greater Portland area. Eligible participants (Table 1) were administered three Ayurveda constitutional questionnaires at two separate time points, one month apart, using REDCap [17], an online survey tool. Participants were able to access the questionnaires online from any computer with Internet access. They were given exactly one week to complete the questionnaires at each time point. All participants provided informed consent.

We studied three constitutional questionnaires used at prominent western Ayurveda educational institutions and made publicly available on their websites. Two questionnaires were developed by the institutions, while the other originated with an established private international Ayurveda products company.

Questionnaire 1 answer options were a degree of agreement on a scale from 0 to 6 (0 = does not apply, 3 = applies somewhat, 6 = applies most); items were grouped into sections labeled as relating to *vata, pitta,* or *kapha dosha*. Questionnaires 2 and 3 gave three distinct answers to each item, corresponding to the three *doshas* and labeled as V, P, and K (and therefore also not disguised).

None of the questionnaires assign a numerical scoring key nor is there an established guideline provided for in previous research.

Table 1
Inclusion/Exclusion criteria.

Participants inclusion criteria	• Adults 18+ years of age • Willingness to complete the three questionnaires online at two separate time points, one month apart • Access to a computer and internet • Ability to read and write in English
Participants exclusion criteria	• Self-reported history or current diagnosis of cognitive impairment that would reduce his or her ability to complete the questionnaires • Anyone who has previously completed an Ayurveda constitutional questionnaire and can recall his or her constitutional type. • Anyone who has previously been given an Ayurveda constitution (such as: *vata, pitta, kapha,* or any combination thereof) by a healthcare professional and can recall his or her constitutional type.

For the purpose of this study, we used the largest of V,P,K scores to determine primary *dosha*. If both remaining (V,P,K) scores were 25% of the total score, the participant was considered to have a single *dosha*. If one of the remaining V,P,K scores was ≥25% of the total score, the participant was considered to have *Dvidosha*, with the second largest score determining secondary *dosha*. If all three scores were ≥25%, the participant was considered to have *Tridosha*. If a participant scored equally *Dvidoshic* at one timepoint (i.e. *vata-pitta/pitta-vata*) and single *dosha* at 2nd time point (i.e. *vata-pitta*) that was at least ½ matching, this was considered *dosha* agreement.

All three questionnaires instructed participants to answer based on what was true for them in general, not based on situations that may have come up in the prior few weeks. This instruction is meant to differentiate between *Prakriti*, which remains stable throughout one's life, and *Vikriti*, which is a temporary state of imbalance. Each questionnaire used its own version of this instruction. The questionnaires were duplicated in REDCap as closely to printed version as possible.

Sample size was estimated by the number of participants needed to detect a mean change of 0.2 (20%) in a single standardized constitution measure (e.g., *vata* score), ranging from 0 to 1. Note that this is less than what would be associated with any change in discrete diagnosis [18]. Since no prior data on variability or test−retest correlations were available, we used very conservative estimates for the variability between individuals (SD = 0.2) and correlation between tests in the same individual (r = 0.3). Using the program G*Power (v.3.1.9.2) [19] with these estimates and a paired t-test design, we calculated a necessary sample size of 17, in order to obtain 90% power to detect a mean change of 0.2 in the constitution score, with a customary alpha level of 0.05. Statistical software, SPSS v.22, was used for statistical analysis.

2.1. Test−retest reliability

In order to determine test−retest reliability of the three individual questionnaires at one month interval, the results for each questionnaire were first standardized as a vector of three numerical values (V, P, K), representing proportions of *vata*, *pitta*, and *kapha* in the *dosha*, summing to 100%. Reliability was tested through a distance measurement (measure of absolute deviation between Test 1 and Test 2 scores), by agreement in diagnosis, and by intra-class correlation co-efficients [20].

Each questionnaire gives separate, summed *vata*, *pitta*, and *kapha* scores. For each questionnaire, standardized proportions were determined by dividing each separate score by the sum so that the total V + P + K was equal to 1 [8]. Results of each questionnaire were compared at two different time points. If the time 1 score is (v_1, p_1, k_1) and the time 2 score is (v_2, p_2, k_2), then the distance D is calculated according to the formula

$$D = \sqrt{(v_1 - v_2)^2 + (p_1 - p_2)^2 + (k_1 - k_2)^2}$$

The composite distance D between Test 1 and Test 2 (V, P, K) vectors varies between 0 and $\sqrt{2}$, with 0 meaning the two vectors are the same and $\sqrt{2}$ meaning that the diagnoses are completely different [6,8]. In addition, to assess distance, intra-class correlations of the test and retest V and P/(P + K) scores were also computed. Note that the V, P, K space is two-dimensional, and that V, P, and K are inversely correlated, by design. P/(P + K) gives a second generator for the space, not *a priori* correlated with V.

As a secondary analysis, we also assigned an Ayurveda constitutional type to each participant. Numerical V,P,K scores were converted to 10 possible Ayurveda diagnoses (or classes). *Dosha* was defined based on a weighting of the three *dosha* types: *vata*, *pitta*, and *kapha* using the rule that any score equal to or above 0.25

was included in the constitution and order would determine primary and secondary *dosha*. None of the questionnaires assigned a numerical scoring key nor was there an established guideline provided in previous research, and thus this weighting system was arbitrarily assigned for this study. For example, suppose an individual's V, P, K scores were 0.6,0.3,0.1. Since two scores- V and P were equal to or above 0.25, a *Dvidosha* constitution was assigned. Furthermore, since V (0.6) was larger than P (0.3), the constitution was *vata-pitta*. Once discrete types were defined, we calculated the proportion of participants for whom the questionnaire produces the same constitutional diagnosis, at both tests.

2.2. Internal consistency

Internal consistency could be tested for Questionnaire 1, because it used designated lists of items for assessing each of the *vata*, *pitta*, and *kapha* constitutional types. A Cronbach's alpha statistic was used to calculate internal consistency from pairwise correlations between items that aim to assess the same *dosha* [12,13].

As an exploratory analysis, we also computed intra-class correlation co-efficients to assess the agreement between the three questionnaires, in assessing prominence of the three *doshas* at baseline.

2.3. Methods by comparison to other self-reporting questionnaire studies

Kurande et al. assessed inter-rater reliability of pulse, tongue, and *Prakriti* assessment through Ayurveda practitioner inspection, history taking, and palpation, and by a researcher developed self-reporting questionnaire comprised of 75 items, 25 items relating to each of the three *dosha* types − *vata*, *pitta*, and *kapha* [8]. Each question of the self-reporting questionnaire required the subject to choose a level of agreement between 1 of 3 options. Study subjects were second-year Ayurveda college students, and none were asked if they had previously been given a *Prakriti* diagnosis. In our study, we excluded participants who had any prior knowledge of their *Prakriti* diagnosis in order to minimize subjective bias. The Kurande et al. questionnaire was administered at 1 time point and diagnosis was compared to other *Prakriti* assessment tools, unlike in our study in which we measured test−retest reliability of each questionnaire to itself over 2 time points, 1-month apart. Statistical analysis for our study was carried out similarly to Kurande et al., weighting each class of diagnosis type and using distance measurements to define level of agreement. In our study, Cronbach's alpha statistic was used for Questionnaire 1 to assess internal consistency.

Rastogi et al. developed a prototype *Prakriti* analysis tool assessing inter-rater reliability by comparing *Prakriti* diagnosis of a self-reporting questionnaire from 1 time point to that of Ayurveda physician diagnosis [9]. As mentioned earlier, we avoided using physician assessment comparison due to not having a validated physician assessment standard by which to compare.

The Rastogi and Kurande et al. studies attempted to minimize environmental or disease factors that may influence external expression of *Prakriti* by selecting healthy volunteers of a young age [8,9]. While our study did not exclude participants with disease conditions, our hope was that measurement of test−retest reliability of each questionnaire from 2 time points would minimize this influence.

Similar to our study, Shilpa et al. developed a *Tridosha*, self-reporting questionnaire tested for reliability over 2 time points, approximately 1 month apart [10]. We chose to test the prior mentioned self-reporting questionnaires vs building upon the work

of Shilpa et al., in order to assess reliability of what is currently used by western trained Ayurveda practitioners.

3. Results

Baseline socio-demographic characteristics of study participants were calculated (Table 2). Of the 26 participants recruited, 19 completed three questionnaires at two time points, one month apart (Fig. 1). Age range was 21–62 years old with a mean age of 34. Of the 19, 5 were male and 14 female. 16 participants were from Portland, Oregon and surrounding areas, and 3 were from out-of-state. 19 participants were included in final analysis of test–retest reliability.

For all questionnaires, the mean (SD) distance between test and retest (V,P,K) was found (Table 3). Intra-class correlation co-effi-cients-ICC (1,1) were calculated for *vata* and the proportion of *pitta* among *pitta* and *kapha* P/(P + K). ICC (1,1) results indicate what proportion of variation in outcomes is due to real differences between individual participants, as opposed to variability of diagnosis in the same individual.

Percentage of test/retest agreement for discrete diagnoses was also calculated using the rule that any score equal to or above 0.25 was included in the constitution and order would determine primary and secondary *dosha* (Table 4). The 3 most common diagnoses across both time points listed in descending order from most to least common was *Tridosha*, *vata-pitta*, and *pitta-vata* tied with *pitta-kapha* (Table 4). The least common diagnosis across both time points was *kapha* tied with *kapha-vata*. Percentage of agreement between time points for each individual questionnaire was also calculated resulting in Q1 with highest level of agreement (Table 4).

Table 2
Baseline characteristics of study participants.

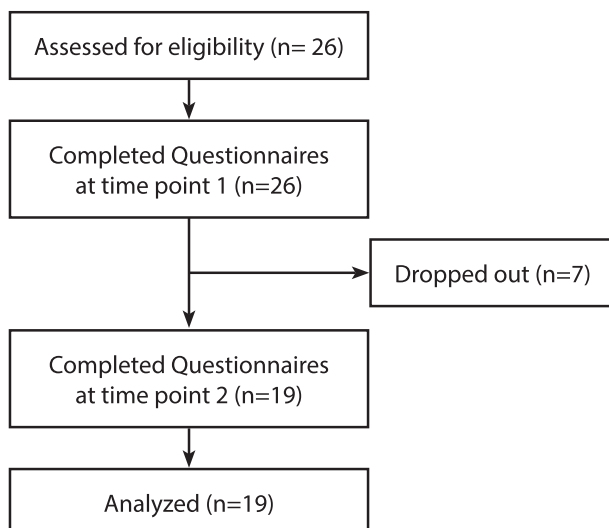| Characteristic	Enrolled N = 26	Completed N = 19
Sociodemographic		
Age, year, mean (SD)	34 (10)	35 (11)
Gender, no. (%)		
Female	18 (69)	14 (74)
Male	8 (31)	5 (26)
Location, no. (%)		
Portland, OR & Surrounding Area	23 (88)	16 (84)
Out-of-State	3 (12)	3 (16)

Fig. 1. Study diagram.

Table 3
Test–retest agreement results (Distance and intraclass correlation coefficients ICC [1,1]).

Questionnaire (Q)	Mean (SD) Distance	Vata ICC (1,1)	95% CI	P/P + K ICC (1,1)	95% CI
Q1	0.05 (.03)	0.69	(0.37, 0.87)	0.69	(0.37, 0.87)
Q2	0.11 (.07)	0.86	(0.67, 0.94)	0.81	(0.58, 0.92)
Q3	0.10 (.07)	0.89	(0.76, 0.96)	0.76	(0.51, 0.91)

Table 4
Test–retest discreet diagnosis agreement results and total.

	Time Point 1 (n)			Time Point 2 (n)			Total
	Q1	Q2[a]	Q3	Q1	Q2	Q3[b]	
Vata	0	2	0	0	3	0	5
Pitta	0	1	1	0	0	0	2
Kapha	0	0	1	0	0	0	1
Vata-Pitta	2	2.5	7	1	3	6.5	22
Vata-Kapha	0	0	0	0	2	0	2
Pitta-Vata	0	1.5	4	0	3	1.5	10
Pitta-Kapha	0	2	2	0	2	5	11
Kapha-Vata	0	1	0	0	0	0	1
Kapha-Pitta	0	0	0	0	2	1	3
Tridosha	17	9	4	18	4	5	57
Total	19	19	19	19	19	19	114
% agreement	Q1: 95% Q2: 42% Q3: 63%						

[a] Q2 at Time Point 1 resulted in 1 participant scoring equally *Dvidoshic* as both *vata-pitta/pitta-vata*, counted as 0.5 for each *dosha*.
[b] Q3 at Time Point 2 resulted in 1 participant scoring equally *Dvidoshic* as both *vata-pitta/pitta-vata*, counted as 0.5 for each *dosha*.

In order to determine a discrete diagnosis from questionnaire scores, it is necessary to choose some minimum cutoff for inclusion of a type in the diagnosis. Since the 0.25 standard was arbitrarily set (the questionnaires themselves provide no guidance), and since this resulted in clinically counterintuitive findings of high numbers of *Tridosha* diagnoses for Q1, we performed a post-hoc re-analysis of discrete diagnoses using a 0.30 standard for constitutional inclusion. The three most common diagnoses across both time points in re-analysis were *vata-pitta*, *pitta-vata*, and *Tridosha*, with *kapha* the least common (Table 5). Percentage of agreement between time points for each individual questionnaire was calculated again in re-

Table 5
Test–retest discreet diagnosis agreement results and total post-hoc re-analysis with 0.30 standard.

	Time Point 1 (n)			Time Point 2 (n)			Total
	Q1[a]	Q2[b]	Q3	Q1	Q2[c]	Q3[d]	
Vata	0	3	1	0	5	2	11
Pitta	0	1	4	0	2	3	10
Kapha	0	1	1	0	0	0	2
Vata-Pitta	4.5	3.5	6	4	3	5.5	26.5
Vata-Kapha	1	0	0	2	1	0	4
Pitta-Vata	2.5	4.5	4	4	2	4.5	21.5
Pitta-Kapha	2	2.5	1	2	2.5	3	13
Kapha-Vata	2	1	0	1	0	0	4
Kapha-Pitta	0	1.5	1	0	2.5	1	6
Tridosha	7	1	1	6	1	0	16
Total	19	19	19	19	19	19	114
% agreement	Q1: 58% Q2: 63% Q3: 63%						

[a] Q1 at Time Point 1 resulted in 1 participant scoring equally *Dvidoshic* as both *vata-pitta/pitta-vata*, counted as 0.5 for each *dosha*.
[b] Q2 at Time Point 1 resulted in 2 participants scoring equally *Dvidoshic* as both *vata-pitta/pitta-vata* and *pitta-kapha/kapha-pitta*, counted as .5 for *dosha*.
[c] Q2 at Time Point 2 resulted in 1 participant scoring equally *Dvidoshic* as both *pitta-kapha/kapha-pitta*, counted as .5 for *dosha*.
[d] Q3 at Time Point 2 resulted in 1 participant scoring equally *Dvidoshic* as both *vata-pitta/pitta-vata*, counted as .5 for *dosha*.

Table 6
Calculation of Cronbach's α, for baseline responses to Questionnaire 1.

	α	Interitem correlations for 14 questions (range)	Internal consistency
Vata	0.523 (N = 26)	−0.33 to 0.62	poor
Pitta	0.604 (N = 25)[a]	−0.45 to 0.76	questionable
Kapha	0.184 (N = 23)[b]	−0.46 to 0.60	unacceptable

[a] When imputing possible values for one missing answer in one participant, α varies from 0.583 to 0.604.
[b] One participant was missing many scores. When imputing possible values for one missing answer in each of two other participants, α varies from 0.206 to 0.373 (N = 25).

Table 7
Intra-class correlation co-efficients-ICC (2,1) for absolute agreement between questionnaires at baseline.

	ICC (2,1)
Vata	0.520
Pitta	0.185
Kapha	0.376

analysis, resulting in a comparable degree of agreement for all three questionnaires (Table 5).

Calculation of Cronbach's α, for baseline responses to Questionnaire 1, produced estimates of moderate to low internal consistency (Table 6), with especially poor consistency in the assessment of *kapha*.

As an exploratory analysis, we calculated intra-class correlation co-efficients for absolute agreement between the three questionnaires at baseline (Table 7). Scores show that there is meaningful lack of agreement between the three scales.

4. Discussion

Many experienced practitioners use some form of Ayurvedic constitutional questionnaires to guide the initial intake, while also taking into account, other tools such as pulse diagnosis, health history, and physical exam (including tongue assessment). Nevertheless, many experienced Ayurveda practitioners do not consider self-reporting constitutional questionnaires to be very reliable as a stand-alone diagnostic tool. This may explain the tolerance of wide variation in questionnaire type and format.

Several studies have examined reliability of *Prakriti* diagnosis by way of pulse taking alone [1,6,18], and by incorporating several of the other diagnostic methods together (pulse + history taking + observation) [1,8,9]. To date, results of these studies have varied greatly from low to moderate levels of reliability. As a result, validity was not considered an objective of this study, due to lack of any validated standard to which questionnaire results can be compared. We recognize this limitation and realize that physician assessment may need to be included for future studies regardless of lacking a validated standard.

None of the questionnaires disguised *dosha* categories in question. Because of this, it was important in the exclusion criteria that participants have not previously received an Ayurveda diagnosis. By the very nature of reading the responses within each labeled category, however, participant answers may have become gradually biased throughout the process of taking each questionnaire, so that by the last one taken, they begin to identify with a specific *dosha*. It would be preferable, in a standardized questionnaire, for items and/ or responses corresponding to different *doshas* to be randomly ordered, without identifiable labels. Furthermore, it would have been beneficial to exclude participants with any known pathology in order reduce influence on external expression of *Prakriti*.

Questionnaire 1 is uniquely different than Questionnaires 2 and 3. The answer options within each section of Questionnaire 1 are a degree of agreement on a scale from 0 to 6 (0 = does not apply, 3 = applies somewhat, 6 = applies most) with each statement pertaining to each particular *dosha-vata*, *pitta*, and *kapha*. Questionnaires 2 and 3 ask participants to choose 1 of 3 possible responses for each question, with each answer corresponding to one *dosha*. Research in scale construction for social and personality psychology tells us that different scale types may produce different responses, even when the question is the same [21]. Response type may influence the accuracy of assessment and therefore, the accuracy of Ayurveda diagnosis. According to Questionnaire 1, but not to Questionnaires 2 and 3, most participants are *Tridoshic*. A *Tridoshic Prakriti* is generally one that is in perfect balance, not requiring treatment, and is infrequently diagnosed [8,18]. This finding could be the function of the 25% cut-off that was arbitrarily assigned for this study in order to assign discrete diagnoses. Although this was not stated as an aim, it is notable in assessing the quality of this questionnaire. Interestingly, post-hoc reanalysis with a 30% cut-off for discrete diagnoses corrected for this high degree of *Tridoshic* results in Q1, but also lowered the rate of agreement between time points for Q1 from 95% to 58%. Clinical application of self-reported questionnaires ultimately relies on diagnosis, and variation in results due to the cut-off standard may impact clinical decision-making associated with treatment.

Cronbach's α statistics showed variable internal consistency in measuring the three *dosha* types, with especially poor consistency in the assessment of *kapha*. The discrepancy in consistency brings up two important questions. 1) Is this questionnaire better adept in evaluating those with *vata* and *pitta* primary *dosha* types and 2) Might inconsistency of *kapha* responses relate to a western cultural bias against the identifiable cluster of *kapha* characteristics?

A *kapha dosha* body type is structurally bigger and of a slower metabolism. Although not portrayed negatively in Ayurveda theory, *kapha dosha* types may be perceived negatively in western culture where thin-framed body types are generally more desirable [22]. Overall, *kapha* was the *dosha* diagnosis least represented amongst participants in this study (Tables 4 and 6). This may be due small sample size (N = 19), difficulty assessing *kapha* through a self-reported questionnaire format, psychologically influenced based on desirability of *dosha* traits, or due to administering the survey mainly to naturopathic students who may not represent a typical distribution of body types.

It is important to note that *Prakriti* features described in questionnaires were originally observed within the context of an Indian population, and thus may not be verbatim transferable to other ethnic groups and geographies. In particular, our sample consisted primarily of white American graduate students, a highly specialized population. Although details have not been published, it is certainly possible that there is better evidence for reliability and validity of the questionnaires, in other populations. While we do not know the background development of specific questionnaires used for this study, we need to take this limitation into consideration for development of any future self-reporting Ayurveda questionnaire.

Within each set of items in Questionnaire 1, we identified a wide range of inter-item correlations, including negative relationships. Future research may include factor analytical techniques aimed at identifying a set of items with high inter-correlation that may be more reasonably identified as markers of the underlying constitutional type. This type of analysis could also be employed with questions having designated *vata*, *pitta*, and *kapha* responses, in order to identify sets of items for which there are high rates of response agreement. Development of such internally consistent item lists will be an important future step in producing a valid questionnaire.

It is interesting to note that some questionnaires demonstrate small mean changes over time (using the distance formula) and have

relatively good intra-class correlations; but that extracted discrete diagnoses still change for most participants. Depending upon how the scales are used, this might greatly impact our assessment of test—retest reliability. Moreover, we should note that this small study was only powered to find meaningful changes in the distance metric [6,8]; our estimates of Cronbach's α, intra-class correlation, and rate of discrete test—retest agreement in *dosha* are relatively imprecise.

5. Conclusion

Analysis performed in this study indicates that three prominent self-reporting Ayurveda questionnaires have moderately good test—retest reliability according to numerical scores, but highly variable reliability according to discrete Ayurveda diagnosis. Internal consistency pertaining to individual constitutions within Questionnaire 1 is at best questionable for all three primary *doshas*, and poor for *kapha*. Further research is necessary to develop a reliable and standardized constitutional questionnaire.

Sources of funding

Conflict of interest

None.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jaim.2017.04.011.

References

[1] 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;4:67—76.

[2] Svoboda RE. Ayurveda, life health and longevity. New Delhi, India: Penguin Books; 1992.

[3] Khalsa KPS, Tierra M. The way of Ayurvedic herbs. Twin Lakes, WI: Lotus Press; 2008.

[4] Svoboda RE. Prakriti. 2nd ed. Twin Lakes, WI: Lotus Press; 1998.

[5] Sharma AK, Kumar R, Mishra A, Gupta R. Problems associated with clinical trials of ayurvedic medicines. Braz J Pharmacogn 2010;20:276—81.

[6] Kurande VH, Waagepetersen R, Toft E, Prasad R. Reliability of pulse diagnosis in traditional indian ayurveda medicine. In: 8th Annual Congress of the International Society for Complementary Medicine Research (ISCMR). Res Complement Med/Forsch Komplementmed, vol. 20(Suppl. 1); 2013. p. 1—9.

[7] Bhalerao S, Patwardhan K. *Prakriti*-based research: good reporting practices. J Ayurveda Integr Med 2016;7(1):69—72.

[8] Kurande V, Bilgrau AE, Waagepetersen R, Toft E, Prasad R. Interrater reliability of diagnostic methods in traditional Indian Ayurvedic medicine. Evid Based Complement Altern Med 2013 Sep. 26;2013:1—12.

[9] Rastogi S. Development and validation of a prototype prakriti analysis tool (PPAT): inferences from a pilot study. Ayu 2012;33:209—18.

[10] Shilpa S, Murthy CG. Development and standardization of mysore tridosha scale. Ayu 2011;32:308—14.

[11] Trochim WMK. Types of Reliability, Research Methods Knowledge Base. http://www.socialresearchmethods.net/kb/reltypes.php [accessed 01.12.13.].

[12] Gliem J, Gliem R. Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales. Midwest research-to-practice conference in adult, continuing, and community education. Columbus, OH: The Ohio State University; October 8-10, 2003 [accessed 13.12.14.], https://scholarworks.iupui.edu/bitstream/handle/1805/344/Gliem+&+Gliem.pdf?sequence=1.

[13] Tavakol M, Dennick R. Making sense of Cronbach's alpha. Int J Med Educ 2011;2:53—5.

[14] Steffen T, Seney M. Test-retest reliability and minimal detectable change on balance and Ambulation tests, the 36-item Short-form health survey, and the Unified Parkinson Disease Rating Scale in People with Parkinsonism. Phys Ther 2008;88:733—46.

[15] Matza L, Thompson C, Krasnow J, Brewster-Jordan J, Zyczynski T, Coyne K. Test-retest reliability of four questionnaires for patients with overactive bladder: the overactive bladder questionnaire (OAB-q), patient perception of bladder condition (PPBC), urgency questionnaire (UQ), and the primary OAB symptom questionnaire (POSQ). Neurourol Urodyn 2005;24:215—25.

[16] Strand LI, Ljunggren AE, Bogen B, Ask T, Johnsen T. The Short-Form McGill Pain Questionnaire as an outcome measure: test—retest reliability and responsiveness to change. Eur J Pain 2007;12:917—25.

[17] REDCap Research Electronic Data Capture. http://project-redcap.org/. [accessed 01.01.14.].

[18] Kurande V, Waagepetersen R, Toft E, Prasad R. Intrarater and interrater reliability of pulse examination in traditional Indian Ayurvedic medicine. Integr Med Res 2013;2:89—98.

[19] Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007;39:175—91.

[20] Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psych Bull 1979;86(2):420—8.

[21] Ackerman RA, Donnellan MB, Roberts BW, Fraley RC. The effect of response format on the psychometric properties of the Narcissistic personality inventory consequences for item meaning and factor structure. Assessment 2015:1—18. 1073191114568113.

[22] Carr D, Jaffe K. The psychological consequences of weight change trajectories: evidence from quantitative and qualitative data. Econ Hum Biol 2012;10(4):419—30.