scientific reports

OPEN

Development of a valid Chinese version of the Cumberland Ankle Instability Tool in Chinese-speaking patients with chronic ankle instability disorders

Wei Wang[1,2✉], Dongfa Liao[1,2], Xia Kang[1,2], Wei Zheng[1✉], Wei Xu[1], Song Chen[1] & Qingyun Xie[1✉]

As an effective scale for the condition assessment of patients with chronic ankle instability (CAI), the Cumberland Ankle Instability Tool (CAIT) is the most widely used scale, and its original version is written in English. Therefore, the purpose of our study is to apply the CAIT to Chinese patients and evaluate its responsiveness, reliability, and validity in terms of Chinese patients with CAI. First, we adapted the CAIT into the Chinese edition (CAIT-C), through which cross-cultural adaptation and translation can be carried out in a five-step procedure. Next, recruited patients completed the three periods of the Foot and Ankle Ability Measure (FAAM), CAIT-C, and the Medical Outcomes Study Short-Form 36 (SF-36) scales. Afterward, to assess the responsiveness, reliability, and validity, we calculated the standardized response mean (SRM), effect size (ES), Spearman's correlation coefficient (r_s), minimal detectable change (MDC), standard error of measurement (SEM), intraclass correlation coefficient (ICC), and Cronbach's alpha. Generally, in the use of CAI, 131, 119, and 86 patients favorably completed the three periods of the scales. The CAIT-C was proven to have good test–retest reliability (ICC = 0.930) and fine internal consistency (Cronbach's alpha = 0.845–0.878). The low-value of MDC (0.04–2.28) and SEM (1.73) show it is possible to detect clinical changes when we take advantage of CAIT-C. Good or moderate correlations (r_s = 0.422–0.738) were gained from the physical subscales of the SF-36 and the subscales of the FAAM and the CAIT-C. Fair or poor correlations (r_s = 0.003–0.360) were gained between the mental subscales of the SF-36 and the CAIT-C, which sufficiently indicated that the CAIT-C had good validity. Moreover, good responsiveness was observed in the CAIT-C (ES = 1.316, SRM = 1.418). The CAIT-C scale is an effective, valid, and reliable tool to evaluate Chinese CAI patients.

The ankle joint is most vulnerable in daily activities as an important weight-bearing joint in the human body. The risk of repeated injury often increases greatly after the first ankle joint sprains[1]. Repeated sprains and instability of the ankle joint are in reciprocal causation, and at least 30% of ankle sprains will develop chronic ankle instability (CAI)[2,3]. The common symptoms of CAI are persistent pain in the ankle joint, repeated sprains, and recurrent "muscle weakness"[4]. Young athletes are more at risk of getting the initial injury[5]. However, chronic CAI impacts a wide age group of people, including those who have quit sports activities entirely[6,7]. According to relevant reports, the number of ankle sprains in the United States is as high as 23,000–27,000 per day[8]. Currently, the prevalence of CAI in the Chinese population has not been reported with high reliability. However, due to the large population of Chinese people, China may have a large number of CAI patients suffering from ankle sprains.

The greater prevalence of CAI and its impact on patients' quality of life has led medical researchers and workers to focus more on diagnosing and treating this disease. Patient-reported outcome measures (PROMs) are among the most important tools for researchers to conduct relevant studies[9]. Since the 1980s, numerous studies have been carried out to develop patient-reported outcome measures (PROMs)[10]. The PROMS customarily

[1]Department of Orthopedics, The General Hospital of Western Theater Command, Tianhui Road 270, Chengdu City 610000, People's Republic of China. [2]These authors contributed equally: Wei Wang, Dongfa Liao and Xia Kang. ✉email: 345880492@qq.com; zw770880@126.com; xqingyun@163.com

collects relevant data in the form of an independent questionnaire. Through these questionnaires, doctors can have more acquaintance with the severity of the patient's state and provide more advisable treatment for the patient[11]. PROMs, which feature high efficiency, low costs, and good reliability, have been a subject of interest in research and clinical practice[12].

The advantages of PROMs mentioned above make them widely applied in various groups of patients. In accordance with the goal, we can divide PROM into specific scales and generic scales. Specific scales can be applied to specific patients. For example, the Western Ontario Shoulder Instability Index (WOSI) is suitable for patients with unstable shoulder joints[13], the Foot and Ankle Ability Measure (FAAM) is appropriate for diverse neuromuscular skeletal changes in the ankle/foot[14], and the Cumberland Ankle Instability Tool (CAIT) is fitting for CAI[15]. The latter is used to assess the sufferers' general state, for instance, the most ordinary Medical Outcomes Study Short-Form 36 (SF-36).

CAIT is one of the most widely used and reliable PROMSs for CAI patients. As recommended by the National Athletic Trainers' Association (NATA) and other professional organizations, PROMs can be used to identify how patients perceive ankle instability, thus helping to make treatment decisions during the management of CAI[16]. Hiller et al. developed the CAIT[15], a discriminative scale used to identify CAI patients and evaluate the severity of functional ankle instability. CAIT is used in various countries worldwide for its ease of use, proven validity, and reliability. The International Ankle Consortium suggests adopting CAIT and other reliable and valid questionnaires to examine ankle instability self-reported by patients[17].

Like most other typical PROMs, CAIT was originally written in English. If there were no language or cultural differences, it could have been used worldwide. When patients from different cultural backgrounds are treated with a reliable and effective scale, it is vital to test the psychometric properties of the scale instead of simply translating content to avoid assessment deviation secondary to cultural differences[18,19]. To apply CAIT to more people with CAI who speak different languages and have different cultural backgrounds, it has been compiled in six different languages (Japanese, Persian, Dutch, Spanish, etc.) by many studies[11,20–25]. Although a previous study complied and translated it into Chinese[26], it lacks an analysis of the validity of the scale, which is the most important and necessary psychometric assessment. In addition, whether the subjects were right for the study has yet to be identified (ordinary people rather than patients with CAI were selected). As a result, we think it is necessary to compile CAIT more accurately and systematically across cultures, translate it into Chinese, and apply it to the largest number of CAI patients[27,28].

Thus, we aimed to translate CAIT into the Chinese Version (CAIT-C) and assess the responsiveness, reliability, and validity of the CAIT-C in CAI patients.

Methods

Translation and cross-cultural adaptation. The principles of previously published guidelines were followed to translate the CAIT from the original version[10,29]. The whole process was composed of five steps. The specific contents have been detailed in a similar article published in our previous publication[30].

Patients and data acquisition. Consecutive native patients who had CAI, spoke Chinese, and visited Chengdu Military General Hospital from February 2016 to March 2018 were enrolled in this study. The inclusion criteria were as follows: (1) age > 18 years with independent signing authority and (2) they reported no less than two cases of severe ankle sprains and a series of feelings including chronic pain, ankle instability, and/or "giving way" in daily life or sports activities. The exclusion criteria were as follows: (1) previous surgical musculoskeletal structures and fractures requiring readjustment in the history of the lower extremity limbs; (2) severe injury to the musculoskeletal structures of the lower limb joints over the past three months; and (3) other chronic inflammatory diseases in the lower limbs that might impact ankle function. Patients who satisfied these criteria and were willing to participate in this study remained under the premise that the sample capacity standard for PROM research was put forward by Terwee et al.[14]. More than one hundred patients' questionnaires were used for internal consistency analysis, and more than fifty patients' questionnaires for ceiling or floor validity, effects, and reliability analysis. All participants read and signed the informed consent form approved by our ethics committee (Chengdu Military General Hospital).

On the first day of admission to the hospital, the patients were required to offer demographic information and, in a quiet meeting room, complete four scales independently. The SF-36, CAIT-C, FAAM, and SC-IdFAI (for another study) were included. One day before the beginning of physiotherapy, which was 1 week after the first set of scales, they completed the CAIT-C for the second time to assess the scale of test–retest reliability. Patients were excluded if they had related treatment in the previous week. Finally, patients who voluntarily received 8 weeks of physiotherapy at our hospital completed the CAIT-C for the third time following therapy to assess responsiveness.

Scales. The CAIT comprises nine items with multiple options related to different aspects of CAI, such as ankle pain, subjective instability during daily and physical activities, and the ankle's response to episodes of giving way[24]. The nine items generate a total score ranging from 0 to 30, with lower scores indicating more severe instability and 30 as the best possible score. The original study established a cutoff score of ≤ 27 to identify those with CAI[15].

The FAAM is a region-specific scale designed to assess the function of the foot and ankle[31]. It consists of two subscales: activities of daily living (ADL) and sports. The ADL subscale and sport subscale score ranges are 0–84 and 0–32, respectively. The higher the score, the better the functional status. The FAAM is a region-specific scale rather than a disease-specific scale; however, it has been proven to have good validity in patients with CAI[32]. The SF-36 is a common quality of life evaluation scale, and 8 subscales of 35 items were included. It can assess a patient's state, including social function, mental health, and physiological function. Each subscale of the SF-36

has its particular marking method, and the ultimate score is changed to a 100-point system. Similarly, the patient's quality of life, or functional status, is increased with increasing scores[33]. The abovementioned scales have been translated into Chinese versions, and it has been proven that these editions are of great responsiveness, reliability, and validity[34,35].

Psychometric assessments and statistical analysis.   Reliability is the degree to which a measurement is free from error[36]. The reliability tests of CAIT-C chiefly contain measurement error, internal consistency, and test–retest reliability. The degree of internal consistency is described as the degree of interaction among projects[35], which is chiefly assessed by the scale of Cronbach's α value of the scale. When α > 0.7, 0.8, and 0.9, the scale has acceptable, good, and excellent internal consistency, respectively[14]. However, extremely high values of Cronbach's α (> 0.95) also show item redundancy[37]. Additionally, Cronbach's α was calculated for the CAIT-C, so if an item was removed, one could see if the item negatively influenced Cronbach's α[14,38]. The test–retest reliability of the scale is assessed in comparison with the previous two responses of patients to CAIT-C. The intraclass correlation coefficient (ICC), rooted in the two-way analysis of the variance in a random effect model, is its assessment indicator. Once ICC > 0.9 and 0.8, the scale has excellent and good reliability, respectively[39]. To observe the systematic errors between surveys, we have further depicted the Bland–Altman diagram[40]. The measurement error includes the randomness and systematic error, which the patient scores, and has nothing to do with the real change of the structure tested[41]. It was calculated in accordance with the formula and analyzed using the standard error of measurement (SEM): $SD \times \sqrt{(1 - ICC)}$. In the first evaluation, the standard deviation of all patients was expressed in SD[41]. The minimal detectable change (MDC) reflects the minimum individual change of fraction, which can be understood as a real change. It was calculated as $SEM \times 1.96 \times \sqrt{2}/\sqrt{n}$ at the group level and $SEM \times 1.96 \times \sqrt{2}$ at an individual level[42].

We can evaluate the validity of the CAIT-C through its construct validity and content validity. The evaluation of the relevance of the items and comprehensiveness is contained in content validity[43]. The three project comprehensive evaluation indexes are patients' feedback, the response rate, and ceiling/floor effects. Assuming that the ceiling/floor effect is lower than 15%, the feedback of the patients is more than 95%, and the patients in the filling scale have no feedback on the difficulty of understanding, then the judgment scale has great comprehensiveness[14,44]. In addition, we invited one rehabilitation specialist and two orthopedic specialists to help judge whether the items were relevant for the construct to be measured and for the patients with CAI[43]. Since the gold standard for assessing CAIT-C standard validity does not exist, the hypothesis test is used to assess the construct validity of CAIT-C. Construct validity is the extent to which the scores on a scale are consistent with hypotheses based on the assumption that the scale validly measures a specific construct[43]. In this study, we selected the FAAM and SF-36 as the control scales for the CAIT-C. Based on the contents of each scale, we assumed that the CAIT-C might have good correlations with the physical subscales of the SF-36 (physical functioning, role physical, bodily pain, general health) and FAAM, and poor correlations with the mental subscales of the SF-36 (vitality, social functioning, role emotional, mental health). In addition, we assumed that the correlations between CAIT-C and FAAM might be stronger than those between CAIT-C and SF-36. Details of the relevant hypotheses among the abovementioned scales are shown in Table 4. Based on the above hypotheses, we calculated Spearman's correlation coefficient (r_s) between the CAIT-C, SF-36, and FAAM by using the results of the patients' first fill-out scale and evaluated the construct validity of the CAIT-C by comparing the consistency between these data and the hypotheses detail. Good construct validity was based on meeting the criterion for at least 75% (8/10 or more) of stipulated a priori hypotheses[41]. The correlations were judged as excellent ($r_s = 0.8$–1.0), good ($r_s = 0.6$–0.8), moderate ($r_s = 0.4$–0.6), fair ($r_s = 0.2$–0.4), or poor ($r_s = 0$–0.2)[30].

Responsiveness is a scale to detect the capability of the structure to be measured over time[43]. We compared the results of the 8-week scale before and after physiotherapy to assess the responsiveness of CAIT-C. The two indicators of reactivity evaluation are standardized response mean (SRM) and effect size (ES). We define the SRM by dividing the average change between each time point by the SD of this change. ES indicated the average change in treatment outcome within 8 weeks before and after the operation, divided by the SD of CAT-C before treatment[45]. When the SRM and ES values exceed 0.80, they are large; when the values are between 0.51 and 0.80, they are intermediate; and when they are less than 0.50, they are small[46].

Statistical Package for the Social Sciences, version 20.0 (SPSS, Chicago, IL, USA), was utilized for statistical analysis. The data are expressed as the mean and standard deviation (SD). The 95% confidence intervals (CIs) can report the ICC value. A P value less than or equal to 0.05 was considered statistically significant.

Ethical statement.   All procedures performed in this study involving human participants were carried out in accordance with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. All participants read and signed informed consent, and this clinical study obtained the approval of the ethics committee of our hospital (The General Hospital of Western Theater Command).

Results
Patients.   A total of 161 patients with CAI (104 males and 57 females) who came to our hospital from February 2016 to March 2018 met the screening criteria. In the end, 132 patients (82% of the invited, 46 women, and 86 men) were invited to participate, and all patients completed the scale. One week later, 119 patients (81 males and 38 females) completed CAIT-C for the second time through reexamination in our hospital, an inquiry by telephone, or emails. Of the 13 patients who did not complete the second questionnaire, nine patients were excluded because they had received the relevant treatment (physical therapy or analgesic drugs) in the previous week, and four patients were out of contact. In addition, 86 of all patients (104 males and 57 females) received regular physicotherapy in our hospital, and they completed CAIT-C for the third time after all treatments were

Characteristics	Number (%) or Mean ± SD
Age (years)	26.5 ± 5.7
Range	18–47
Age groups	
≦ 20	23 (17.4%)
21–30	81 (61.4%)
31–40	25 (18.9%)
≧ 41	3 (2.3%)
Gender	
Female	46 (34.8%)
Male	86 (65.2%)
Affected side	
Right	98 (74.2%)
Left	34 (25.8%)
Bilateral	
BMI (kg/m²)	23.4 ± 4.9

Table 1. Demographic and clinical characteristics of participants. This study and the manuscript with DOI number "https://doi.org/10.1186/s12891-020-03314-1" published by the same research team used the same cohort of Chinese CAI patients. *BMI* body mass index.

Items	Corrected item: total correlation[a]	Cronbach's α if item was deleted
Overall scale	1.000	0.873
Item 1	0.762	0.858
Item 2	0.809	0.847
Item 3	0.756	0.854
Item 4	0.758	0.853
Item 5	0.710	0.861
Item 6	0.727	0.857
Item 7	0.826	0.845
Item 8	0.503	0.877
Item 9	0.537	0.878

Table 2. The internal consistency of CAIT-C. *CAIT-C* Chinese version of Cumberland Ankle Instability Tool. [a]Calculated by the Spearman's correlation coefficient of the items with total score.

completed (8 weeks later). Thus, to evaluate the validity of the CAIT-C, measurement error, retest reliability, and internal consistency, we selected 132 samples, of which 119 samples were used to evaluate the retest reliability of the CAIT-C, and 86 cases were used to evaluate CAIT-C reactivity. Table 1 shows the detailed demographic data of the primary participants.

Translation and cross-culture adaptation process. The translation of CAIT, forward and backward, was very smooth. Since it was easy to comprehend the items of CAIT, we had not improved them. Twenty patients (10 women and 10 men) completed the final version of the CAIT-C in CAI patients. No patient indicated that the project was difficult to understand or that it lacked standardization.

Reliability. The Cronbach's α for the CAIT-C was 0.873, providing good internal consistency. Moreover, suppose that the Cronbach's α coefficient of each item was deleted, as shown in Table 2, the correlation coefficient between each item's score and the remaining total score. In the analysis of the project, no improvement was found every time the items were deleted from the scale, except for items 8 and 9. When omission was omitted, the project increased slightly.

The ICC value of CAIT-C was 0.930, indicating that CAIT-C had excellent test–retest reliability (Table 3). In addition, Bland–Altman plots showed no systemic error in the first two rounds (Fig. 1), which confirmed that CAIT-C had a good test–retest agreement.

The SEM value of CAIT-C was 1.73. Therefore, the MDC reflecting the minimal individual and group (this study) change in score that can be interpreted as a real change was 4.80 and 0.44.

Validity. In this research, there was no error in response to the CAIT-C questionnaire. The distribution of scores indicated there was no floor effect (1.5%) or ceiling effect (3.8%) in the CAIT-C (Table 3). In addition,

	Current study	Dutch version	Persian version	Korean version	Portuguese version	Spanish version	Original version
Floor effect[a]	1.5%	1%	2.6%	–	0%	0%	–
Ceiling effect[a]	3.8%	2%	5.1%	–	7.9%	9%	–
ICC (CI range)	0.930 (0.901–0.951)	0.943 (–)	0.91–0.95 (0.80–0.97)	0.94 (–)	0.95 (0.93–0.97)	0.95–0.98 (0.93–0.99)	0.96 (–)
SEM	1.73	0.82	2.03–2.40	1.72	–	–	–
MDC (I)[b]	4.80	2.28	5.6–6.5	–	–	–	–
MDC (G)[c]	0.44	0.04	–	–	–	–	–
ES	1.316	–	–		0.75	0.69–1.07	–
SRM	1.418	–	–		–	–	–

Table 3. The floor/ceiling effects, test–retest reliability, measurement error and responsiveness of CAIT-C. *CAIT-C* Chinese version of Cumberland Ankle Instability Tool, *ICC* intraclass correlation coefficient, *CI* confidence interval, *SEM* standard error of measurement, *MDC* minimal detectable change, *ES* effect size, *SRM* standardized response mean. [a]Percentage of patients with the worst (floor effect) and the best (ceiling effect) score. [b]The MDC value at an individual level. [c]The MDC value at the group level.
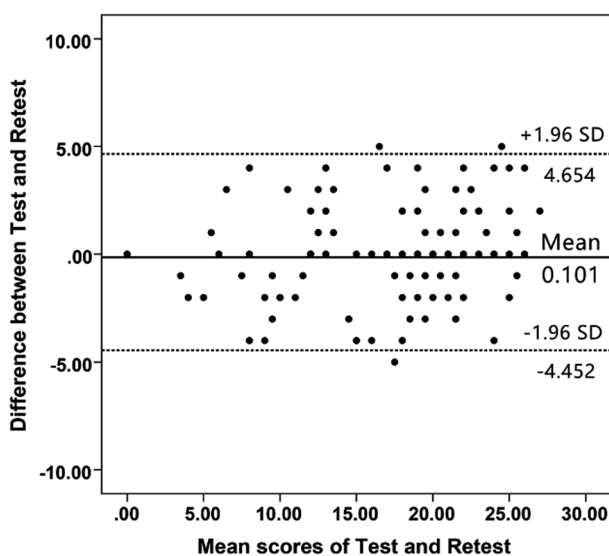
Figure 1. Bland–Altman plots of the test–retest reliability of the CAIT-C. Each data point indicates how the difference between the two test sessions for an individual patient compares to the mean of the two sessions for scores of each CAIT-C. The interval of two sessions was 1 week. The dashed line shows the 95% (± 1.96 SD) limits of agreement.

no patient indicated that the contents of CAIT-C were difficult to understand. According to the assessment and analysis of two departments of orthopedics experts and rehabilitation experts, the amount of information obtained from each CAIT-C project is sufficient to assess the health-related quality of life of CAI patients. Hence, it is not recommended to remove or add any items. According to the above results, the CAIT-C has good content validity.

Table 4 shows the relevant data of the CAIT-C construct validity evaluation. It showed good (r_s = 0.624 to 0.738) correlations between the two subscales of the FAAM and CAIT-C, moderate (r_s = 0.422–0.560) correlations between the physical subscales of the SF-36 and CAIT-C, and fair or poor (r_s = 0.080–0.260) correlations between the mental subscales of the SF-36 and CAIT-C. The above results were completely consistent with our a priori hypotheses (10/10).

Responsiveness. The questionnaires were completed before and after physicotherapy to assess the responsiveness of CAIT-C, and the relevant data is listed in Table 3. Overall, the average CAIT-C score increased after treatment. The values of SRM (1.418) and ES (1.316) are both greater than 1, which suggests that CAIT-C has good responsiveness.

Scales	Spearman's correlation coefficient (r_s) [a]	P value	Hypotheses
FAAM			
ADL	0.738	< 0.0001	≥ Moderate, and better than SF-36 wih CAIT-C
Sport	0.642	< 0.0001	
SF-36			
Physical function	0.443	< 0.0001	≥ Moderate, and worse than FAAM wih CAIT-C
Role-physical	0.560	< 0.0001	
Bodily pain	0.522	< 0.0001	
General health	0.422	< 0.0001	
Vitality	0.260	0.003	≤ Poor, and worse than physical subscales of SF-36 and CAIT-C
Social function	0.140	0.109	
Role-emotional	0.080	0.360	
Mental health	0.183	0.036	

Table 4. Construct validity of the CAIT-C. *CAIT-C* Chinese version of Cumberland Ankle Instability Tool, *FAAM* foot and ankle ability measure, *ADL* activity of daily living, *SF-36* Short-Form 36. [a]Calculated by the Spearman's correlation coefficient (r_s) of the CAIT-C with FAAM and SF-36.

Discussion

In clinical surveys, PROMs are tools of great importance. Researchers can compare the questionnaire reports from similar studies and quantify the functional condition of patients. This is very helpful for an increasing number of multicenter clinical studies[10]. Today, in China, clinical research is developing rapidly, and many relevant papers are published every year. This is because there is a large number of patient groups in China and also because of the government's emphasis on scientific research[47]. Therefore, China is now in great need of effective PROMs. These scales can help many patients in China receive a more accurate diagnosis and treatment and provide support for many clinical studies in China.

CAIT is one of the most widely used PROMs for CAI patients. Only one study has reportedly performed the cross-cultural translation of CAIT in Chinese[26], but the validity indexes of CAIT-C, the most important part, were not evaluated in that study. There was also no assessment of measurement error in the reliability analysis or determination of whether or not the study sample (ordinary people rather than patients with CAI were selected) used to assess reliability was appropriate. Another study on CAIT-C's cutoff scores was reported, but it does not involve the Psychometric Assessments of CAIT-C[48]. Therefore, it is necessary to conduct a more accurate and comprehensive study on the cross-cultural compilation of CAIT in Chinese.

Before discussing the results, the limitations of this study deserve attention. First, China's population may not be fully represented because the sample size is limited. Second, considering translation, the language we use is simplified Chinese, as is the official language. However, as a multiethnic country, many ethnic groups in China have their own languages, such as Hong Kong, Macao, Xinjiang, and the Tibetan Plateau. Hence, in the survey, ethnic cultural differences deserve attention. Last, there was some loss of participants due to exclusion criteria and loss of follow-up, but the overall sample appears to be adequately powered based on the results.

The process of intercultural adaptation and translation is relatively smooth in this study. We believe that the original version of the CAIT project is suitable for the Chinese cultural background. Therefore, we have not adapted the content of questions, which may also benefit from the easy-to-understand advantages of CAIT[21].

The CAIT-C had good internal consistency (Cronbach's α = 0.845–0.878), and its Cronbach's α was slightly higher than that in the original version and other language versions[12,15,21–25]. Simultaneously, we found that Cronbach's α of CAIT-C would be slightly higher (0.877 and 0.878) when item 8 or item 9 was removed, which also appeared in the Korean version and the Persian version[20,22]. The correlations between the scores of the two items and the total score were the weakest (r_s = 0.503–0.537). This might be because item 8 and item 9 were set in the hypothesis context ("roll over on ankle"), while other items were about the daily life of the patient, which caused differences in the same patient responding to these items. Good test–retest reliability (ICC = 0.930) is reflected in the CAIT-C, which is consistent with the results of similar studies (Table 3). Additionally, we consider that the assessment of the CAIT-C test–retest reliability as more appropriate using a week as the time interval because the patient is less prone to forget the specific answers in the previous questionnaire within a week, and the patient's functional status and daily life would not dramatically change in 1 week. MDC and low values for measurement error mean that small clinical changes and individual-level changes can be detected at the population level by CAIT-C.

There was no floor effect or ceiling effect in CAIT-C. The evaluation of three experts also authenticated that the CAIT-C items were well correlated with the patient's prognosis and CAI patients. In addition, due to the easy-to-understand advantage of CAIT-C, there were no missed responses in any returned questionnaires. Based on the above objective results, and the good feedback from patients who filled out the questionnaire, the CAIT-C had good content validity.

In other cross-cultural adaptation studies on CAIT, except for the Dutch version[21], the remaining versions all evaluated the criterion validity of CAIT[20,22–25]. However, in light of the COSMIN list (consensus-based Standards for the selection of health status Measurement INstruments), which is a consistency-based checklist for assessing the methodological quality of the measurement attributes of the health measurement instruments based on an

international Delphi study, the "criterion validity" was defined as the degree of a PROMs instrument that reflects the degree of "gold standard", in 2010[43]. The standard used should be reasonably considered the "gold standard", but the Delphy group agreed that there was no gold standard for PROMS instruments[43]. The "hypotheses testing" for evaluating the so-called "criterion validity" in other cross-culture adaptation studies was the method for assessing the construct validity of CAIT. By hypothesis testing, the correlations between the CAIT-C and the subscales of the SF-36 and FAAM in this study were the same as that of our previous hypothesis, meaning that the CAIT-C has good construct validity. The CAIT-C had the strongest correlations with the two subscales of the FAAM. Although the FAAM is not a disease-specific scale for patients with CAI, it mainly concerns the functional status of the patient's foot and ankle (region-specific scale), such as CAIT. Therefore, the objective of FAAM items is very close to that of CAIT. In addition, the CAIT-C had weak correlations with the mental subscales of the SF-36, but the correlation still existed ($P < 0.05$), indicating that the functional status of the foot and ankle in CAI patients would affect their psychological states.

One of the important factors in determining whether the scale can be used in prospective clinical research is the quality of the scale's responsiveness. In this study, CAIT-C showed good responsiveness, which means that CAIT-C can be sensitive to changes in the functional condition of patients after systemic physicotherapy. Compared with related studies, the ES value of this study was slightly higher (ES = 0.69–1.07)[12,24,25]. This might be because patients in this study received 8 weeks of physicotherapy, and the treatment period in other studies was shorter (3–4 weeks), which led to certain differences in the degree of improvement in the patient's functional status.

Conclusions

In summary, we successfully translated CAIT into Chinese. After verification, the version was easy to use and has good responsiveness, reliability, and validity. Hence, we advise that CAIT-C be used in assessing the functional condition of Chinese CAI patients in related clinical work or clinical studies to help researchers or doctors collect the necessary data.

References

1. Ajis, A., Younger, A. S. & Maffulli, N. Anatomic repair for chronic lateral ankle instability. *Foot Ankle Clin.* **11**(3), 539–545 (2006).
2. Tanen, L., Docherty, C. L., Van Der Pol, B., Simon, J. & Schrader, J. Prevalence of chronic ankle instability in high school and division I athletes. *Foot Ankle Spec.* **7**(1), 37–44 (2014).
3. van Rijn, R. M. *et al.* What is the clinical course of acute ankle sprains? A systematic literature review. *Am. J. Med.* **121**(4), 324–331 (2008).
4. Hertel, J. Functional anatomy, pathomechanics, and pathophysiology of lateral ankle instability. *J. Athl. Train.* **37**(4), 364–375 (2002).
5. Nelson, A. J., Collins, C. L., Yard, E. E., Fields, S. K. & Comstock, R. D. Ankle injuries among United States High School sport athletes, 2005–2006. *J. Athl. Train.* **42**, 381–387 (2007).
6. Hiller, C. E., Kilbreath, S. L. & Refshauge, K. M. Chronic ankle instability: evolution of the model. *J. Athl. Train.* **46**, 133–141 (2011).
7. Hiller, C. E. *et al.* Prevalence and impact of chronic musculoskeletal ankle disorders in the community. *Arch. Phys. Med. Rehabil.* **93**, 1801–1807 (2012).
8. Verhagen, R. A., de Keizer, G. & van Dijk, C. N. Long-term followup of inversion trauma of the ankle. *Arch. Orthop. Trauma Surg.* **114**(2), 92–96 (1995).
9. Hill, C. L., Parsons, J., Taylor, A. & Leach, G. Health related quality of life in a population sample with arthritis. *J. Rheumatol.* **26**, 2029–2035 (1999).
10. Guillemin, F., Bombardier, C. & Beaton, D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J. Clin. Epidemiol.* **46**, 1417–1432 (1993).
11. Guyatt, G. H., Feeny, D. H. & Patrick, D. L. Measuring healthrelated quality of life. *Ann. Intern. Med.* **118**(8), 622–629 (1993).
12. De Noronha, M., Refshauge, K. M., Kilbreath, S. L. & Figueiredo, V. G. Cross-cultural adaptation of the Brazilian-Portuguese version of the Cumberland Ankle Instability Tool (CAIT). *Disabil. Rehabil.* **30**(26), 1959–1965 (2008).
13. Kirkley, A., Griffin, S., McLintock, H. & Ng, L. The development and evaluation of a disease-specific quality of life measurement tool for shoulder instability. The Western Ontario Shoulder Instability Index (WOSI). *Am. J. Sports Med.* **26**, 764–772 (1998).
14. Martin, R. L., Irrgang, J. J., Burdett, R. G., Conti, S. F. & Van Swearingen, J. M. Evidence of validity for the Foot and Ankle Ability Measure (FAAM). *Foot Ankle Int.* **26**, 968–983 (2005).
15. Hiller, C. E., Refshauge, K. M., Bundy, A. C., Herbert, R. D. & Kilbreath, S. L. The Cumberland ankle instability tool: a report of validity and reliability testing. *Arch. Phys. Med. Rehab.* **87**(9), 1235–1241 (2006).
16. Yeung, M. S., Chan, K. M., So, C. H. & Yuan, W. Y. An epidemiological survey on ankle sprain. *Br. J. Sport Med.* **28**(2), 112–116 (1994).
17. Gribble, P. A. *et al.* Selection criteria for patients with chronic ankle instability in controlled research: a position statement of the International Ankle Consortium. *J. Athl. Train.* **49**, 121–127 (2014).
18. Pynsent, P. B. Choosing an outcome measure. *J. Bone Jt. Surg. Br.* **83**, 792–794 (2001).
19. Zheng, W., Li, J., Zhao, J., Liu, D. & Weidong, Xu. Development of a valid simplified Chinese version of the oxford hip score in patients with hip osteoarthritis. *Clin. Orthop. Relat. Res.* **472**, 1545–1551 (2014).
20. Ko, J., Rosen, A. B. & Brown, C. N. Cross-cultural adaptation and validation of the Korean version of the Cumberland ankle instability tool. *Int. J. Sports Phys. Ther.* **10**(7), 1007–1014 (2015).
21. Vuurberg, G., Kluit, L. & van Dijk, C. N. The Cumberland Ankle Instability Tool (CAIT) in the Dutch population with and without complaints of ankle instability. *Knee Surg. Sports Traumatol. Arthrosc.* **26**(3), 882–891 (2018).
22. Hadadi, M. *et al.* Cross-cultural adaptation, reliability, and validity of the Persian version of the Cumberland Ankle Instability Tool. *Disabil. Rehabil.* **39**(16), 1644–1649 (2017).
23. Kunugi, S. *et al.* Cross-cultural adaptation, reliability, and validity of the Japanese version of the Cumberland ankle instability tool. *Disabil. Rehabil.* **39**(1), 50–58 (2017).
24. Rodríguez-Fernández, Á. L., Rebollo-Roldán, J., Jiménez-Rejano, J. J. & Güeita-Rodríguez, J. Psychometric properties of the Spanish version of the Cumberland Ankle Instability Tool. *Disabil. Rehabil.* **37**(20), 1888–1894 (2015).

25. Cruz-Díaz, D., Hita-Contreras, F., Lomas-Vega, R., Osuna-Pérez, M. C. & Martínez-Amat, A. Cross-cultural adaptation and validation of the Spanish version of the Cumberland Ankle Instability Tool (CAIT): an instrument to assess unilateral chronic ankle instability. *Clin. Rheumatol.* **32**(1), 91–98 (2013).
26. Shu-yuan, Li., Lei, Z., Ya-nan, Z. & Zeng-tao, W. On the validity and reliability of the Chinese version of Cumberland ankle instability Tool. *Chin. J. Sports Med.* **30**, 814–819 (2011).
27. Fennig, C. D. *et al.* (eds) *Ethnologue: Languages of Asia* 18th edn. (SIL International, 2015).
28. Lewis, M. P. *Ethnologue: Languages of the World* 16th edn. (SIL International, 2009).
29. Beaton, D. E., Bombardier, C., Guillemin, F. & Ferraz, M. B. Guidelines for the process of cross-cultural adaptation of selfreport measures. *Spine* **25**, 3186–3191 (2000).
30. Terwee, C. B. *et al.* Quality criteria were proposed for measurement properties of health status questionnaires. *J. Clin. Epidemiol.* **60**, 34–42 (2007).
31. Carcia, C. R., Martin, R. L. & Drouin, J. M. Validity of the Foot and Ankle Ability Measure in athletes with chronic ankle instability. *J. Athl. Train.* **43**(2), 179–183 (2008).
32. Brazier, J. E. *et al.* Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *BMJ* **305**, 160–164 (1992).
33. González-Sánchez, M., Li, G. Z., Ruiz Muñoz, M. & Cuesta-Vargas, A. I. Foot and ankle ability measure to measure functional limitations in patients with foot and ankle disorders: a Chinese cross-cultural adaptation and validation. *Disabil. Rehabil.* **39**(21), 2182–2189 (2017).
34. Li, L., Wang, H. M. & Shen, Y. Chinese SF-36 Health Survey: translation, cultural adaptation, validation, and normalisation. *J. Epidemiol. Community Health* **57**, 259–263 (2003).
35. Baumann, F. *et al.* Validation of a German version of the International Hip Outcome Tool 12 (iHOT12) according to the COSMIN checklist. *Health Qual. Life Outcomes* **14**, 3 (2016).
36. Eechaute, C., Vaes, P., Van Aerschot, L., Asman, S. & Duquet, W. The clinimetric qualities of patient-assessed instruments for measuring chronic ankle instability: a systematic review. *BMC Musculoskelet. Disord.* **8**, 6 (2007).
37. Lohr, K. N. *et al.* Evaluating quality-of-life and health status instruments: development of scientific review criteria. *Clin. Ther.* **18**(5), 979–992 (1996).
38. Landis, J. R. & Koch, G. G. The measurement of observer agreement for categorical data. *Biometrics* **33**(1), 159–174 (1977).
39. Bland, J. M. & Altman, D. G. Measuring agreement in method comparison studies. *Stat. Methods Med. Res.* **8**, 135–160 (1999).
40. Tak, I. *et al.* The Dutch hip and groin outcome score: cross-cultural adaptation and validation according to the COSMIN checklist. *J. Orthop. Sports Phys. Ther.* **48**(4), 299–306 (2018).
41. Weir, J. P. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J. Strength Cond. Res.* **19**(1), 231–240 (2005).
42. Mokkink, L. B. *et al.* The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med. Res. Methodol.* **10**, 22 (2010).
43. 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.* **14**, 331–345 (1995).
44. Wang, W. *et al.* Cross-cultural translation of the Lysholm knee score in Chinese and its validation in patients with anterior cruciate ligament injury. *BMC Musculoskelet. Disord.* **17**(1), 436 (2016).
45. Husted, J. A., Cook, R. J., Farewell, V. T. & Gladman, D. Methods for assessing responsiveness: a critical review and recommendations. *J. Clin. Epidemiol.* **53**, 459–468 (2000).
46. Cohen, J. *Applied Multiple Regression/correlation Analysis for the Behavioral Sciences* (Lawrence Erlbaum Associates, 1983).
47. Wei, X. *et al.* Development of a simplified Chinese version of the hip disability and osteoarthritis outcome score (HOOS): cross-cultural adaptation and psychometric evaluation. *Osteoarthr. Cartil.* **20**, 1563–1567 (2012).
48. Li, Y. *et al.* Applicability of cutoff scores of Chinese Cumberland Ankle Instability Tool and Foot and Ankle Ability Measure as inclusion criteria for study of chronic ankle instability in Chinese individuals. *Phys. Ther. Sport* **48**, 116–120 (2020).

Acknowledgements

Author contributions

W.W., D.L., X.K., W.Z. and Q.X. made substantial contributions to this article (conception and design, acquisition of data, analysis and interpretation of data). W.X. has been involved in drafting the manuscript and analysis of data. S.C. has been involved in acquisition of data.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to W.W., W.Z. or Q.X.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.