



Reliability of an adapted core strength endurance test battery in individuals with axial spondylarthritis

Anne-Kathrin Rausch^{1,2} · Philipp Baltisberger¹ · André Meichtry¹ · Beatrice Topalidis³ · Adrian Ciurea⁴ · Theodora P. M. Vliet Vlieland² · Karin Niedermann¹

Received: 10 July 2020 / Revised: 2 September 2020 / Accepted: 15 September 2020 / Published online: 21 September 2020
© The Author(s) 2020

Abstract

Objectives To adapt the core strength endurance test battery (aCSE), previously used for testing athletes, to a target group of patients with axial spondylarthritis (axSpA), to evaluate its intra-tester reliability and its associations with disease-specific factors.

Methods A cross-sectional study was conducted at axSpA exercise therapy groups, including both axSpA patients and the physiotherapist group leaders (PTs). The aCSE was used to measure the isometric strength endurance of the ventral, lateral, and dorsal core muscle chains (measured in seconds), as well as to assess the disease-specific factors of functional status, self-reported pain, and perceived strength performance. The aCSE was repeated after 7–14 days to measure intra-tester reliability for the same rater (PT group leader). Reliability was calculated as an intra-class correlation coefficient (ICC) using a nested design. The associations between ventral, lateral, and dorsal strength endurance and the disease-specific factors were calculated using Pearson correlation coefficients.

Results Study participants were 13 PT group leaders and 62 axSpA patients. The latter were all capable of performing the aCSE, with the exception of one individual. A moderate to substantial intra-rater reliability (ICCs (95%CI)) was found for the ventral (0.54 (0.35, 0.74)), lateral (0.52 (0.33, 0.70)), and dorsal (0.71 (0.58, 0.86)) core muscle chains. None of the aCSE measures correlated with the disease-specific factors.

Conclusion The aCSE was found to be a reliable test battery for assessing core strength endurance in axSpA patients. Interestingly, aCSE performance was not associated with any disease-specific factors.

Key Points

- The adapted core strength endurance test battery measures the isometric strength of the ventral, lateral and dorsal core muscle chains.
- The adapted core strength endurance test battery showed a moderate to substantial intra-rater reliability for all three muscle chains tested in axSpA patients.
- No correlations were found between the adapted core strength endurance test battery and the disease-specific factors of self-reported pain, functional status and perceived strength performance.

Keywords Ankylosing spondylitis · Assessment · Group exercise · Physical therapy

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10067-020-05408-6>) contains supplementary material, which is available to authorized users.

✉ Anne-Kathrin Rausch
rauh@zhaw.ch

¹ School of Health Professions, Institute of Physiotherapy, Zurich University of Applied Sciences, 8401 Winterthur, Switzerland

² Department of Orthopaedics, Rehabilitation and Physical Therapy, Leiden University Medical Center, Leiden, Netherlands

³ Ankylosing Spondylitis Association of Switzerland, 8050 Zurich, Switzerland

⁴ Department of Rheumatology, Zurich University Hospital, 8091 Zurich, Switzerland

Introduction

Axial spondylarthritis (axSpA) is a chronic inflammatory rheumatic condition that mainly affects the axial skeleton, iliosacral joints, and spine. It can lead to structural and functional impairments and can have a significant impact on an individual's quality of life and working ability [1, 2].

The cornerstones of non-pharmacological treatment are patient education and regular exercise [1]. Current recommendations for physical activity in individuals with inflammatory arthritis emphasize that the general physical activity recommendations, comprising the four exercise domains of cardiorespiratory fitness, muscle strength, flexibility, and neuromotor performance, are effective, safe, and feasible for axSpA patients [3].

In addition to peripheral muscle strength, core muscle strength (sometimes referred to as the “powerhouse”) is especially important for people with axSpA [4], since inflammation and reduction in mobility affect the dynamic stabilization of the spine. A certain level of core stability is needed, in terms of strength and muscle fatigue resistance, for the activities of daily living (improved posture, enhanced balance, and proprioception) and sports performance (such as correct and safe barbell-based exercising) [5]. Core muscle performance is a complex and multivariable construct, with core stability not yet having been clearly defined [6, 7]. Experts agreed in a Delphi-project that core stability is “the ability to achieve and sustain control of the trunk region at rest and during precise movement” and depends on the components of muscle strength and neuromotor control [8]. Three common categories to assess core muscle strength can be differentiated according to this concept: maximal strength, power [9, 10], and strength endurance, the latter defined as the ability to sustain a given level of force production over time [11].

No randomized controlled studies have been published to date investigating the effect of muscle strength exercises alone on lower limb and trunk muscle strength performance in axSpA patients [12]. Training the trunk muscles does seem, however, to have a positive effect on flexibility in healthy people [13]. Moreover, the combination of core strength and flexibility exercises has been found to have a positive effect on disease activity and flexibility in axSpA patients (assessed by the Bath Ankylosing Spondylitis Mobility Index, BASMI) [14].

The physical activity recommendations for individuals with inflammatory arthritis [3] underline the fact that regular evaluation of physical fitness, including all aforementioned four domains, should be part of standard care for axSpA patients [3]. The assessment of strength, however, is challenging. Dagfinrud et al. found that the monitoring of muscle strength parameters in group exercise therapy (GET) studies was essentially nonexistent [15]. A survey of the axSpA GET in Switzerland and the Netherlands confirmed this finding for

clinical practice [16]. However, there is a need to regularly assess all fitness dimensions to identify those with the potential for improvement. This could help healthcare providers and patients in the process of promoting physical activity, e.g., giving advice, setting goals, planning, and performance of exercises.

The Ankylosing Spondylitis Association of Switzerland (Schweizerische Vereinigung Morbus Bechterew, SVMB) currently has more than 4000 members and organizes more than 60 exercise groups across Switzerland. Participants exercise weekly in groups led by physiotherapists (PTs) in land-based or water-based settings. The objectives of the GET are to minimize the progressive spinal mobility restriction [17], reduce cardiovascular and biological risk factors [18], and maintain or increase muscle strength. According to the recommendations [3], annual fitness assessments are also part of the SVMB's quality concept. In clinical settings, such as the axSpA GET, the fitness dimensions should be assessed by easy-to-use and inexpensive devices. In this target group, strength assessment should focus on core muscle strength endurance rather than peripheral muscle strength.

In the absence of an existing gold standard, the isometric core strength endurance test battery (CSE), originally developed by the Swiss Olympic Medical Centers for use with athletes, seems to be the only available tool for use in an axSpA GET setting. It was found to be easy to perform, inexpensive to use, and to present with good psychometric properties [19]. The Biering-Sorensen test, which is part of the aCSE, showed good to excellent inter-rater and test-retest reliability [20], but no prior data on intra-rater reliability were found. The CSE was designed to evaluate the “basic” core strength of athletes, meaning the minimum strength required for the performance of sports [19]. With the aid of a reference table including norm data of Swiss athletes, assessors judge whether core strength is sufficient or insufficient.

The aim of this study was to adapt the CSE for axSpA patients (aCSE), evaluate intra-tester reliability and the associations with disease-specific factors.

Methods

Design

For the analysis of reliability, a cross-sectional study with a nested design was conducted, given that across the groups, a different numbers of individuals were measured by the PTs (raters). The findings are reported in line with the GRRAS (Guidelines for Reporting Reliability and Agreement Studies) guidelines [21].

Participants

Physiotherapists (PT)

The PTs leading the GET in the German-speaking region of Switzerland ($n = 45$) were invited to participate in this study and to perform and rate the aCSE test battery with their group participants. They were requested to attend a 2-h practical training session (lead by PB). In addition, they received detailed step-by-step explanations in the form of videos, photos, and handouts. Thirteen PTs registered for the aCSE training.

AxSpA patients

An information letter was sent to the 206 participants of the SVMB exercise groups of the 13 PTs registered for the aCSE training. The letter explained the purpose and procedures of the study and described the aCSE test battery. Inclusion criteria were age of over 18 years and sufficient German language skills. Eligible individuals also had to be capable of getting down onto the floor and lying in the prone, supine, and side positions. Exclusion criteria were heart diseases of class three or four, according to the New York Heart Association [22], diagnosed osteoporosis of grade two or more [23], or surgical spondylodesis performed on the entire lumbar spine.

Assessments

The aCSE test battery was performed by each axSpA patient twice with a period of 1 to 2 weeks between the tests (T0 and T1). A period of 1 week allowed for full recovery of the muscles after performing the first aCSE test battery at T0, while the maximum of 2 weeks minimized the possible effect of disease-related changes. The measurements took place during the regular GET. At T0, as well as recording patient characteristics, measurements of pain status, disease activity functional restriction, and physical fitness were made. The same group PT leader conducted the aCSE measurements at both time points with each patient. No encouragement (in terms of cheering) was allowed.

Participant characteristics

Gender, age, disease duration, and self-reported exercise hours per week were recorded for axSpA patients. Gender, age, and work experience in years were documented for the PTs.

Core strength measurements

Adaptation and performance of the core strength endurance test battery (aCSE)

Given the large number of exercise groups that could potentially implement the aCSE, the criteria for test selection were validity, reliability, and low cost. A further important criterion was feasibility, i.e., the test had to be able to be performed in sports halls with little specific equipment and by individuals with a broad range of fitness and health.

Previous field testing had shown that the CSE used with athletes [19] was too demanding for some SVMB exercise group participants. They reported joint pain while in positions of lateral and ventral muscle chains. The following adaptations were consequently made to the CSE for testing axSpA patients, resulting in an adapted version named the aCSE:

- The starting position for the ventral plane was changed from plank to quadruped position.
- Tests were performed statically instead of dynamically [24] to ensure a constant load (this modus reduced the risk of injury and was easier to standardize).
- Rods were used for standardization, i.e., the participant was asked to keep contact with the horizontal rod during the test (see supplement, Fig. 1).

The time that the subject was able to remain in each of three positions was measured in seconds. Time recording stopped whenever the participant lost contact with the rod for the third time. The participant could get into the required position for a maximum of three attempts. Before the next position was measured, the participant rested for at least 30 s. Detailed instructions on the performance of aCSE can be found in the [Supplement](#).

Additional outcomes and assessments

Pain intensity

Pain intensity was measured using the Numeric Rating Scale (NRS), which is acknowledged as a reliable and valid measurement tool in clinical practice and research [25]. The NRS measures pain on a scale from 0 to 10 (0 = no pain at all and 10 = worst pain ever). In this study, the tool was used to assess average pain immediately before and after the aCSE test batteries were performed.

The Bath Ankylosing Disease Activity Index (BASDAI)

AxSpA disease activity was measured using the BASDAI, which is a valid and reliable self-reporting questionnaire [26]. The questionnaire consists of six items to determine pain

in the peripheral joints and spine, fatigue, morning stiffness, and pain sensitivity to touch [27]. The BADAI results in a mean score of 0–6 points (0 = no disease activity, 6 = highest disease activity) [27].

The Bath Ankylosing Spondylitis Functional Index (BASFI)

Functional limitations experienced by axSpA patients during ten everyday tasks were evaluated using the BASFI, which is a valid and reliable self-reporting questionnaire [26]. The BASFI evaluation results in a mean score between 0 and 10 (0 = no handicap and 10 = highest possible degree of functional limitation).

Physical fitness questionnaire (FFB-Mot.)

Self-perceived motor performance ability during everyday tasks, with regard to cardiorespiratory fitness, strength, flexibility, and coordination, was evaluated using The Physical Fitness Questionnaire (FFB-Mot., German: Fragebogen zur Erfassung des motorischen Funktionsstatus) [28]. It has been shown that the FFB-Mot. is a valid instrument to determine physical fitness in a healthy adult population [29]. The outcome of the 28 items on the questionnaire results in a total score ranging from 5 to 140; the strength questions include seven items for which the score can range from 0 to 30. For example, a total score of 35/140 for a 58-year-old healthy man would indicate a rather poor overall physical fitness status, while a score of 100/140 would indicate a rather good physical fitness status [29]. It has not yet been used as an assessment for evaluation, so no reliability data are available.

Weekly exercise

Questions on weekly hours of planned exercise, together with the focus of the exercise dimension (muscle strength, cardiovascular, neuromotor, or flexibility), were asked by the PT, together with the patient characteristics (age, gender, year of diagnosis, and disease duration).

Sample size and statistical analysis

Sample size was estimated based on the precision of the reliability estimate following Giraudeau and Mary [30]. With a targeted width of the confidence interval (two margin of errors or four standard errors) of $w = 0.3$, 57 subjects were needed for (a conservative) ICC = 0.65. For larger ICCs, the required sample size would be smaller or the precision in the analysis would be larger.

All statistical analyses were performed using the R statistical software R version 3.5.3 [31].

Reliability

Generalizability theory [32] was used as the framework to estimate the reliability of the raters' time-keeping while assessing core muscle strength. To estimate the intra-rater reliability, a linear mixed model for “(strength endurance) Time” Y of Subject i nested in Rater j measured at Time point k was fitted to the data:

$$Y_{i(j),k} = \mu + S(R)_i + R_j + T_k + RT_{ik} + \epsilon_{i,j,k}$$

with $Y_{i(j),k}$ representing strength endurance time, μ representing the global mean, $S(R)_i$ corresponding to $S_i + S(R)_{ij}$ which cannot be disentangled in a nested design, and ϵ_{ijk} the independent and normal distributed errors.

The intra-rater reliability was calculated as the intra-class correlation coefficient

$$\text{corr} \left(Y_{i(j),k}, Y_{i(j),k'} \right) = \frac{\sigma_{S(R)}^2}{\sigma_{S(R)}^2 + \sigma_T^2 + \sigma_{RT}^2 + \sigma_\epsilon^2},$$

with the σ^2 representing the corresponding variance components. Bootstrapped confidence intervals for the intra-class correlation coefficient (ICC) were computed. Values less than 0.2 demonstrate a slight reliability, values between 0.2 and 0.39 indicate a fair reliability, values between 0.4 and 0.59 describe moderate reliability, values between 0.6 and 0.79 indicate substantial reliability, and values greater than 0.80 indicate almost perfect reliability between measurements [33]. The lower limit of the 95% confidence interval of the ICC to estimate the level of reliability was used [34].

Associations between the aCSE outcomes and disease-specific factors

To evaluate the relationship between aCSE outcomes and patient-reported strength and the disease-specific outcomes, the time measurements for aCSE ventral, lateral, and dorsal were correlated with the FFB-Mot. subscales of strength, NRS-measured pain, BASDAI, and BASFI, using the Pearson correlation coefficient. The size of a correlation coefficient can be interpreted as negligible ($\rho < 0.3$), low ($0.3 < \rho < 0.5$), moderate ($0.5 < \rho < 0.7$), high ($0.7 < \rho < 0.9$), and very high ($\rho > 0.9$), with both positive/negative correlations [35]. A priori, it was hypothesized that the following would provide evidence of an association between aCSE performance and disease-related factors: (a) a positive Pearson correlation (r_s) > 0.3 between aCSE and FFB-Mot. subscale strength and/or self-reported hours exercise per week; and/or (b) a negative correlation < 0.5 between aCSE and pain (NRS) and/or disease activity (BASFI, BASDAI). The lower limits of 95% confidence intervals adjusted for multiple testing were used.

Table 1 Participants' characteristics

Characteristics of axSpA patients	
	<i>n</i> = 62
Gender, female, <i>n</i> (%)	29 (46)
Age, years	54.6 [11.2]
BASDAI (0–10)	3.5 [2]
BASFI (0–10)	1.8 [1.5]
FFB-Mot. total (0–140)	85.8 [18]
FFB subscale strength (0–35)	21.5 [5.4]
NRS pre aCSE at T0	2.2 [2.2]
NRS post aCSE at T0	1.6 [2.4]
Disease symptoms since, years	27.5 [12.9]
Disease duration since diagnosis, years	20.1 [13]
Self-reported training per week, hours	3.1 [2.9]
Characteristics of physiotherapists	
	<i>n</i> = 13
Gender, female, <i>n</i> (%)	9 (69)
Age, years	46 [11]
Work experience, years	6.3 [3.5]
Number of participants per rater (median, range)	4 [2–9]

Data are mean (standard deviation), if not stated otherwise. *BASDAI*, Bath Ankylosing Disease Activity Index; *BASFI*, Bath Ankylosing Spondylitis Functional Index; *FFB-Mot.*, physical fitness questionnaire; *NRS*, numeric rating scale; *aCSE*, adapted core strength endurance test battery; *T0* and *T1*, time point 1 (baseline) and 2, respectively; *GET*, group exercise therapy

Results

Thirteen PTs (28.8%) and 62 group participants (30.0%) provided informed consent and were included in the study; three of the latter could not participate at T1 (*n* = 2 due to acute influenza, *n* = 1 due to pain after the first measurement). Descriptive data for the study participants are shown in Table 1.

Intra-rater reliability of aCSE

A moderate to substantial intra-rater reliability was found for all three test positions that tested the strength of the ventral, lateral, and dorsal core muscle chains (Table 2).

Table 2 Intra-rater reliability of the aCSEs

Test position	ICC	95% CI
aCSE ventral	0.54	0.35, 0.74
aCSE lateral	0.52	0.33, 0.70
aCSE dorsal	0.71	0.58, 0.86

aCSE, adapted core strength endurance test battery; *CI*, confidence interval; *ICC* intra-class correlation coefficient

Table 3 Evaluation of associations between aCSE and disease-specific factors

Association	<i>r</i>	95%CI*
aCSE _v vs. FFB-Mot. strength	0.25	−0.19, 0.61
aCSE _l vs. FFB-Mot. strength	0.34	−0.11, 0.67
aCSE _d vs. FFB-Mot. strength	0.30	−0.15, 0.65
aCSE _v vs. self-reported training hrs	−0.06	−0.44, 0.34
aCSE _l vs. self-reported training hrs	−0.06	−0.44, 0.34
aCSE _d vs. self-reported training hrs	−0.01	−0.32, 0.30
aCSE _v vs. pain	−0.11	−0.50, 0.31
aCSE _l vs. pain	−0.07	−0.44, 0.33
aCSE _d vs. pain	0.06	−0.33, 0.43
aCSE _v vs. BASFI	−0.03	−0.39, 0.33
aCSE _l vs. BASFI	−0.16	−0.54, 0.27
aCSE _d vs. BASFI	−0.35	−0.68, 0.10
aCSE _v vs. BASDAI	−0.23	−0.59, 0.22
aCSE _l vs. BASDAI	−0.26	−0.62, 0.19
aCSE _d vs. BASDAI	0.03	−0.33, 0.37

r, observed correlation coefficient; *CI*, confidence interval (*uncertainty adjusted for multiple testing (Holm procedure)); *BASDAI*, Bath Ankylosing Disease Activity Index; *BASFI*, Bath Ankylosing Spondylitis Functional Index; *FFB-Mot.* strength, physical fitness questionnaire subscale strength; *NRS*, numeric rating scale; *aCSE*, adapted core strength endurance test battery; *v*, ventral; *l*, lateral; *d*, dorsal

Associations of aCSE outcomes with disease-specific factors

Data shown in Table 3 indicate no positive or negative associations between aCSE outcomes and any disease-specific factors measured. Thus, the a priori hypotheses were not confirmed.

Discussion

The aCSE values for the intra-rater reliability of the ventral, lateral, and dorsal planes show a moderate to substantial level of agreement. The results show that PTs are able to use the aCSE reliably, even with little experience of aCSE testing. Furthermore, nearly all participants were able to perform the aCSE, and only one (1.6%) was unable to. In conclusion, the aCSE is a reliable assessment for people with axSpA in a group setting.

Possible associations between aCSE performance and disease-specific factors, according to the a priori hypotheses, were not confirmed. This is an interesting finding, suggesting that the aCSE can be performed by an individual with axSpA irrespective of their perceived strength performance, functional status, and self-reported pain. This may be an aspect worthy of further investigation.

However, we were unable to confirm that pain did not influence aCSE performance. Pain was reported using NRS prior to and post-testing, but no statistically significant negative relationship was found that would imply that more pain is related to less strength. Other than pain [36], the factors of motivation and effort are determinants of strength performance [37]. Midgley and colleagues hypothesized that verbal encouragement could make a difference to test outcomes, through having a positive impact on intrinsic motivation and physical performance [38]. The rating of perceived exertion together with consistent verbal encouragement should be incorporated into the future use of the aCSE. Consistent verbal encouragement could be usefully included in the GET assessment situation.

Due to the complexity of the core strength construct, there has been no gold standard for assessing core strength in the past, or core strength endurance in particular. Core strength is vital for maintaining an upright posture and is especially important for axSpA patients, who are affected by spinal inflammation and decreased spinal mobility. It is therefore appropriate to assess core strength endurance for axSpA patients. The physical position for assessing these individuals poses an additional challenge. It could be argued that to assess strength endurance of the dorsal muscle chain, the testing position described by Ito and colleague [39] (lying in prone position performing back extension) would be preferable to the testing position described by Biering Sorensen [40] (lying in prone position half of the body fixed on a treatment bench holding neutral position), because the Biering Sorensen occasionally activates more hip extensor muscles [24, 41]. However, since the restricted flexibility of the spine is a major impediment in axSpA patients, the test procedure described by Biering Sorensen was considered to be more appropriate for covering the complete range of impairments from low to high. In future, it would be useful to develop a score for strength that encompasses all three planes, or to investigate whether there is a score from one plane that would serve as a sufficient proxy for all three planes. Ultimately, a database containing norm core strength values of individuals both with and without axSpA could ensure a meaningful interpretation of test results.

The original CSE is used as part of a sports performance test battery for athletes [19]. With the aid of a reference table containing norm data of Swiss athletes, assessors judge whether core strength is sufficient or insufficient. Currently, no reference table including norm data of people with axSpA is available. This aspect should be taken into consideration when reporting test results to participants. However, until such a set of norm data is established, intra-individual comparisons might give an orientation.

Surprisingly, no relationship between aCSE outcomes and perceived strength performance was found, although there is evidence that people of all ages have a moderately accurate perception of their physical fitness [42]. In contrast, other

studies [43] provide evidence that people tend to be unrealistic in assessing their abilities, such as physical fitness or level of physical activity. A further reason for our finding could be that the aCSE measures core stability and not strength endurance. Core stability was not explicitly assessed in the patient-reported outcome measure. However, there is a consensus that core stability is dependent on the components of muscle strength and neuromotor control [8]. Majewski-Schrage and colleagues [8] asked experts to provide assessment techniques that were specific to the components of core stability. The top three answers were timed side bridge, Sorensen test, and timed prone bridge. All three of these are elements of the aCSE.

Low-cost was prioritized in our study setting. However, future projects could investigate the use of appropriate, objective, strength assessment devices, such as a hand-held dynamometer [44]. This has been used previously to measure core strength indirectly (albeit maximal strength, not strength endurance) and has been found to be reliable and valid [7, 45]. It might be less time-consuming to incorporate, e.g., easy to perform handgrip, instead of core strength, especially in a group setting.

This study has strengths and limitations. To our knowledge, this was the first study to evaluate the psychometric properties of a core strength assessment in axSpA patients. Movement assessment by visual observation and time-keeping appears to be feasible for PTs. However, it is a limitation of this study that reliability was tested by the clinically working usual GET PT leaders rather than by raters in a laboratory setting. Although the 2-h assessment training was designed to standardize the test procedures and was appreciated by the PTs, their performance was likely to have been influenced by their motivation and understanding of the importance of performing the aCSE in a standardized way. PTs were not asked to perform the additional, disease-specific assessments, such as the Bath Ankylosing Mobility Index (BASMI), since the participants needed to receive specific instructions on how to hold the correct positions and to exert themselves. Additionally, the PTs emphasized that it was challenging to perform the assessments with each individual in the group during the usual GET sessions, in terms of time management and the planned exercise program. These factors (i.e., PTs' adherence to standardization, PTs' motivation, complexity of time-consuming tests) may have had an impact on the accuracy and reliability of the measurement outcomes. The reliability values could potentially be improved with improved test conditions.

Conclusion

Regular fitness assessment on an individual basis is part of the SVMB's concept of quality and is recommended [3]. The

choice of assessment used in this study was influenced mainly by feasibility criteria, such as low-cost, easy-to-use for axSpA patients, and the ability for people with different health and fitness status to perform in both land-based and water-based GET settings. Appropriate assessments with good psychometric quality are necessary. In summary, the aCSE was found to be a feasible assessment instrument to measure core strength in axSpA patients reliably. It could be potentially combined with other assessments of aerobic fitness, flexibility, and neuromotor performance to establish an inexpensive and practical assessment battery, covering all exercise dimensions, for axSpA patients. Future research should establish a reference norm data table including axSpA patients, to enable an appropriate interpretation of test results. In addition, a less time-consuming alternative to implementing aCSE testing should be investigated, e.g., using only one representative plane instead of three separate planes.

Acknowledgments We thank all participating physiotherapists and participants of the SVMB exercise groups for supporting the project with great enthusiasm and effort. Furthermore, our grateful thanks go to René Bräm, CEO SVMB, for promoting the project.

Data of availability Please contact the authors for reasonable data requests.

Authors' contributions PB, AKR and KN designed the project. PB organized the data collection as part of his master's thesis. KN supervised the progress of the study. AM conducted the statistical analysis. All authors contributed significantly to the interpretation of results. AR drafted the manuscript. All authors provided comments on each draft and approved the final version.

Funding Open access funding provided by ZHAW Zurich University of Applied Sciences. The measurement equipment was funded by physioswiss in 2017.

Compliance with ethical standards

Disclosures None.

Ethical approval and consent to participate The study obtained ethical approval from the Ethics committee of Canton Zurich (BASEC nr. 2017-00833). All participants (patients and PTs) gave written informed consent.

Consent for publication Not applicable

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will

need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- van der Heijde D, Ramiro S, Landewe R, Baraliakos X, Van den Bosch F, Sepriano A et al (2017) 2016 update of the ASAS-EULAR management recommendations for axial spondyloarthritis. *Ann Rheum Dis* 76(6):978–991
- Mangone M, Paoloni M, Procopio S, Venditto T, Zucchi B, Santilli V, Paolucci T, Agostini F, Bernetti A (2020) Sagittal spinal alignment in patients with ankylosing spondylitis by rasterstereographic back shape analysis: an observational retrospective study. *Eur J Phys Rehabil Med* 56(2):191–196
- Rausch Osthoff AK, Niedermann K, Braun J, Adams J, Brodin N, Dagfinrud H, Duruoz T, Esbensen BA, Günther KP, Hurkmans E, Juhl CB, Kennedy N, Kiltz U, Knittle K, Nurmohamed M, Pais S, Severijns G, Swinnen TW, Pitsillidou IA, Warburton L, Yankov Z, Vliet Vlieland TPM (2018) 2018 EULAR recommendations for physical activity in people with inflammatory arthritis and osteoarthritis. *Ann Rheum Dis* 77(9):1251–1260
- Fayh A, Brodt GA, Souza C, Loss JF (2018) Pilates instruction affects stability and muscle recruitment during the long stretch exercise. *J Bodyw Mov Ther* 22(2):471–475
- Borghuis J, Hof AL, Lemmink KA (2008) The importance of sensory-motor control in providing core stability: implications for measurement and training. *Sports Med* 38(11):893–916
- Hibbs AE, Thompson KG, French D, Wrigley A, Spears I (2008) Optimizing performance by improving core stability and core strength. *Sports Med* 38(12):995–1008
- Huyghe M, Sercu H, Vervaeke O (2016) Evaluating core stability: reliability and validity of hand-held dynamometry in measuring core strength. Ghent University
- Majewski-Schrage T, Evans TA, Ragan B (2014) Development of a core-stability model: a Delphi approach. *J Sport Rehabil* 23(2):95–106
- Zemkova E, Poor O, Jelen M (2019) Between-side differences in trunk rotational power in athletes trained in asymmetric sports. *J Back Musculoskelet Rehabil* 32(4):529–537
- Sapega AA, Drillings G (1983) The definition and assessment of muscular power. *J Orthop Sports Phys Ther* 5(1):7–9
- Strand SL, Hjelm J, Shoepe TC, Fajardo MA (2014) Norms for an isometric muscle endurance test. *J Hum Kinet* 40:93–102
- Rausch Osthoff AK, Juhl CB, Knittle K, Dagfinrud H, Hurkmans E, Braun J et al (2018) Effects of exercise and physical activity promotion: meta-analysis informing the 2018 EULAR recommendations for physical activity in people with rheumatoid arthritis, spondyloarthritis and hip/knee osteoarthritis. *RMD Open*
- Kim K-J (2010) Effects of core muscle strengthening training on flexibility, muscular strength and driver shot performance in female professional golfers. *Int J Appl Sports Sci* 22(1)
- Fang HY, Cai WZ, Pan YH, Wu DC, Liang LM (2016) Six-month home-based exercise and supervised training in patients with ankylosing spondylitis. *Int J Clin Exp Med* 9(3):6635–6641
- Dagfinrud H, Halvorsen S, Vøllestad NK, Niedermann K, Kvien TK, Hagen KB (2011) Exercise programs in trials for patients with ankylosing spondylitis: do they really have the potential for effectiveness? *Arthritis Care Res* 63(4):597–603
- Rausch Osthoff AK, van der Giesen F, Meichtry A, Walker B, Van Gaalen F, Goekoop-Ruiterman YPM, et al (2019) The perspective of people with axial spondyloarthritis regarding physiotherapy:

- room for the implementation of a more active approach. *Rheumatol Adv Pract*
17. Van den Berg R, Baraliakos X, Braun J, van der Heijde D (2012) First update of the current evidence for the management of ankylosing spondylitis with non-pharmacological treatment and non-biologic drugs: a systematic literature review for the ASAS/EULAR management recommendations in ankylosing spondylitis. *Rheumatology*. 51(8):1388–1396
 18. Sveaas SH, Bilberg A, Berg IJ, Provan SA, Rollefstad S, Semb AG et al (2019) High intensity exercise for 3 months reduces disease activity in axial spondyloarthritis (axSpA): a multicentre randomised trial of 100 patients. *Br J Sports Med*
 19. Bourban PM, Tschopp M. *Manual Leistungsdiagnostik*. In: Swiss Olympic Medical Centers BfS (ed). https://www.swissolympic.ch/dam/jcr:b15b191a-eb0d-46e8-b9c0-417b887a440d/Leistungsdiagnostik_Manual_160201_DE.pdf2016. Accessed 19 June 2017
 20. AbilityLab SR (2015) Rehabilitation measures: Beiring Sorensen Test. <https://www.sralab.org/rehabilitation-measures/beiring-sorensen-test> [11.03.2020]
 21. Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hrobjartsson A et al (2011) Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *Int J Nurs Stud* 48(6):661–671
 22. Committee NYHAC, Association NYH (1979) Nomenclature and criteria for diagnosis of diseases of the heart and great vessels: Little, Brown Medical Division
 23. Kiltz U, Rudwaleit M, Sieper J, Braun J (2016) Evidenzbasierte empfehlung zur diagnostik und therapie der axialen spondyloarthritis. *Z Rheumatol*:1–6
 24. Latimer J, Maher CG, Refshauge K, Colaco I (1999) The reliability and validity of the Biering-Sorensen test in asymptomatic subjects and subjects reporting current or previous nonspecific low back pain. *Spine (Phila Pa 1976)* 24(20):2085–2089 discussion 90
 25. Williamson A, Hoggart B (2005) Pain: a review of three commonly used pain rating scales. *J Clin Nurs* 14(7):798–804
 26. Moncur C (2003) Ankylosing spondylitis measures: The Ankylosing Spondylitis Quality of Life (ASQOL) Scale, Bath Ankylosing Spondylitis Disease Activity Index (BASDAI), Bath Ankylosing Spondylitis Functional Index (BASFI), Bath Ankylosing Spondylitis Global Score (BAS-G), Bath Ankylosing Spondylitis Metrology Index (BASMI), Dougados Functional Index (DFI), Health Assessment Questionnaire for the Spondyloarthropathies (HAQ-S), and Revised Leeds Disability Questionnaire (RLDQ). *Arthritis Care Res* 49(S5):S197–S209
 27. Brandt J, Westhoff G, Rudwaleit M, Listing J, Zink A, Braun J, Sieper J (2003) Validierung einer deutschen Version des Fragebogens BASDAI zur Messung der Krankheitsaktivität bei ankylosierender Spondylitis. *Z Rheumatol* 62(3):264–273
 28. Bös K (2001) *Handbuch Motorische Tests*, 2.: Hogrefe
 29. Bos K, Abel T, Woll A, Niemann S, Tittlbach S, Schott N (2002) The physical fitness questionnaire (FFB-mot). *Diagnostica* 48(2): 101–111
 30. Giraudeau B, Mary JY (2001) Planning a reproducibility study: how many subjects and how many replicates per subject for an expected width of the 95 per cent confidence interval of the intraclass correlation coefficient. *Stat Med* 20(21):3205–3214
 31. R Core Team (2018) *Language and environment for statistical computing*. In: *Computing RfFS* (ed). Vienna, Austria
 32. Brennan RL (2001) *Generalizability theory*. New York
 33. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics*. 33(1):159–174
 34. Koo TK, Li MY (2016) A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 15(2):155–163
 35. Hinkle DE, Wiersma W, Jurs SG (2003) *Applied statistics for the behavioral sciences*. Edition t. Houghtin Mifflin, Boston
 36. Riddle DL, Stratford PW (2011) Impact of pain reported during isometric quadriceps muscle strength testing in people with knee pain: data from the osteoarthritis initiative. *Phys Ther* 91(10):1478–1489
 37. Marcora SM, Staiano W (2010) The limit to exercise tolerance in humans: mind over muscle? *Eur J Appl Physiol* 109(4):763–770
 38. Midgley AW, Marchant DC, Levy AR (2018) A call to action towards an evidence-based approach to using verbal encouragement during maximal exercise testing. *Clin Physiol Imaging* 38: 547–553
 39. Ito T, Shirado O, Suzuki H, Takahashi M, Kaneda K, Strax TE (1996) Lumbar trunk muscle endurance testing: an inexpensive alternative to a machine for evaluation. *Arch Phys Med Rehabil* 77(1):75–79
 40. Biering-Sorensen F (1984) Physical measurements as risk indicators for low-back trouble over a one-year period. *Spine (Phila Pa 1976)* 9(2):106–119
 41. Latikka P, Battié MC, Videman T, Gibbons LE Correlations of isokinetic and psychophysical back lift and static back extensor endurance tests in men. *Clin Biomech* 10(6):325–330
 42. Germain JL, Hausenblas HA (2006) The relationship between perceived and actual physical fitness: a meta-analysis. *J Appl Sport Psychol* 18(4)
 43. Yu CA, Rouse PC, Veldhuijzen Van Zanten JJ, Ntoumanis N, Kitas GD, Duda JL et al (2015) Subjective and objective levels of physical activity and their association with cardiorespiratory fitness in rheumatoid arthritis patients. *Arthritis Res Ther* 17:59
 44. Stark T, Walker B, Phillips JK, Fejer R, Beck R (2011) Hand-held dynamometry correlation with the gold standard isokinetic dynamometry: a systematic review. *PM R* 3(5):472–479
 45. Harding AT, Weeks BK, Horan SA, Little A, Watson SL, Beck BR (2017) Validity and test-retest reliability of a novel simple back extensor muscle strength test. *SAGE Open Med* 5: 2050312116688842

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