The Use of Cluster Analysis by Partitioning around Medoids (PAM) to Examine the Heterogeneity of Patients with Low Back Pain within Subgroups of The Treatment Based Classification System

Esmaeil Shokri[1], Mohsen Razeghi[2,3]*, Hadi Raeisi Shahraki[4], Reza Jalli[5], Alireza Motealleh[2,3]

[1]PhD Candidate, Department of Physiotherapy, School of Rehabilitation Sciences, Shiraz University of Medical Sciences, Shiraz, Iran
[2]PhD, Department of Physiotherapy, School of Rehabilitation Sciences, Shiraz University of Medical Sciences, Shiraz, Iran
[3]PhD, Rehabilitation Sciences Research Center, Shiraz University of Medical Sciences, Shiraz, Iran
[4]PhD, Department of Epidemiology and Biostatistics, Faculty of Health, Shahrekord University of Medical Sciences, Shahrekord, Iran
[5]MD, Department of Radiology, Namazee Hospital, Shiraz University of Medical Sciences, Shiraz, Iran

ABSTRACT

Background: Current evidence in low back pain (LBP) focuses on population averages and traditional multivariate analyses to find the significant difference between variables. Such a focus actively obscured the heterogeneity and increased errors. Cluster analysis (CA) addresses the mentioned shortcomings by calculating the degree of similarity among the relevant variables of the different objects.

Objective: This study aims to evaluate the agreement between the treatment-based classification (TBC) system and the equivalent 3 cluster typology created by partitioning around medoids (PAM) analysis.

Material and Methods: In this cross-sectional study, a convenient sample of 90 patients with low back pain (50 males and 40 females) aged 20 to 65 years was included in the study. The patients were selected based on the 21 criteria of 2007 TBC system. An equivalent 3 cluster typology (C3) was applied using PAM method. Cohen's Kappa was run to determine if there was agreement between the TBC system and the equivalent C3 typology.

Results: PAM analysis revealed the evidence of clustering for a C3 cluster typology with average Silhouette widths of 0.12. Cohen's Kappa revealed fair agreement between the TBC system and C3 cluster typology (Percent of agreement 61%, Kappa=0.36, p <.001). Selected criteria by PAM analysis were different with original TBC system.

Conclusion: Higher probability of chance agreement was observed between two classification methods. Significant inhomogeneity was observed in subgroups of the 2007 TBC system.

Keywords

Treatment based Classification; Low Back Pain; Reproducibility of Results; Cluster Analysis; PAM Analysis

*Corresponding author:
Mohsen Razeghi
School of Rehabilitation Sciences, Shiraz University of Medical Sciences, Abiverdi s1 Street, Chamran Blvd., Shiraz, Iran
E-mail: razeghm@sums.ac.ir

Received: 8 January 2020
Accepted: 31 March 2021

Introduction

In applied research studies, the focus is on population averages and traditional multivariate analysis to find the significant difference between variables [1]. Such a focus actively obscured the heterogeneity between subjects and considered it as a noise [2]. Misleading this heterogeneity can increase errors in estimating the effectiveness of in-

terventions and also in the results of decision analytic models [3]. In many applied research studies, the most useful level of analysis involves the consideration of the entire case more than comparisons between variables within each case [1]. Cluster analysis (CA) is the process of grouping of seemingly homogenous objects into different subgroups [4]. Recently partitioning around mediods (PAM) analysis was suggested to discover and identify unknown patterns without any limitation.

Some evidence disclosed the heterogeneity in patients with low back pain (LBP) [5, 6]. Several systems, such as "Mechanical Diagnosis and Therapy" [7], "Treatment-based Classification" [8], "Movement System Impairment" [9], and the "O'sullivan Classification System" [10] have been developed for classification of patients with LBP that all of these systems have focused on the traditional multivariate analyses for subgrouping in their methods [8,11-13]. A few studies also tried to directly uncover the heterogeneity through CA [14-17]. However, they only involved psychosocial and/or work related aspects of LBP and neglected physical examinations and biomechanical dimensions.

Among the different classification approaches, the treatment-based classification (TBC) system has received the highest priority in research [18]. This system firstly developed in 1995, and subsequent clinical trials have shown that clinical decision-making based on this system provides better outcome than no classification in patients with LBP [8, 19, 20], and has revised and updated many times based on clinical experience and scientific evidence [21, 22]. The developers of TBC algorithm emphasized that it should be revised to optimize its comprehensiveness, refine current criteria and explore additional treatments [23]. This system, like other classification approaches, does not cover all patients with LBP and 25% of patients remained unclassified [23]. Some previous studies have reported moderate interobserver reliability for this system [23-26].

Since all criteria in the TBC system were examined individually and separately, the behavior of one criterion in the presence or absence of the others may affect the final outcome of classification. The mentioned shortcomings were not addressed previously. Therefore, the aim of present study was to evaluate the reliability of TBC system using equivalent C3 cluster typology by PAM analysis.

Material and Methods

In this cross-sectional study, a convenience sample of 118 patients with low back problems was screened. Based on 8 primary criteria of the TBC system and considering 5 patients for each criterion included in the clustering, the initial sample size was calculated as 40 patients. Afterward, the initial clustering was performed to identify the number of criteria that was important in clustering. After initial clustering, the maximum number of possible influential criteria was determined 18 patients and the final sample size according to rule of thumb was determined as 90 patients.

Patients in this study were 50 males and 40 females aged 20 to 65 years, who met the eligibility criteria. Informed consent was obtained from all subjects before participation. The patients in the LBP group were selected if they reported pain or discomfort localized between T12 to the gluteal fold with or without referral pain to the legs [27]. In order to identify the patient, who candidate for rehabilitation treatment, TBC system (2015 version) was used to meet the first level of triage [22]. Patients with red flags of serious pathology (e.g. pathologic fractures, sacral stress fracture, acute spondylolisthesis, cancer, etc.) or serious comorbidities that do not respond to the standard rehabilitation management (e.g. rheumatoid arthritis, central sensitization) were excluded from the study [22]. The patients were also excluded if they were currently pregnant or if they had a positive history of back surgery in the preceding 6 months, history of using spinal fusion, scoliosis rod or screws, history of tak-

ing steroid medications in the last month or if they received physiotherapy at least 3 months prior to their entrance. We also excluded the patients, classified as self-care management based on 2015 TBC system. Therefore, if the patients had the Oxford depression questionnaire (ODQ) score of less than 6% [28, 29] or pain score of less than 4 mm, were excluded from study [22].

Physical examinations and 2007 TBC criteria: All patients completed the Persian version of Oswestry disability questionnaire (ODI) for low back disability [30] and their pain was assessed by visual analogue pain scale [31]. They also completed the Persian version of Fear-Avoidance Beliefs Questionnaire (FABQ) to assess psychological aspects of their work and physical activity [32]. Since the TBC system (2015 version) [22] has not yet been completely developed, the TBC system (2007 version) [21] was used for classification in this study. Evaluation was done by a physical therapist familiar with the TBC system.

Neurological examinations, including muscle strength, sensation, reflexes and the result of straight-leg-raise (SLR) test were recorded [21]. The range of motion (ROM) for SLR test was measured using an inclinometer in both sides [33]. Spinal movements were evaluated for presence of any aberrant motion. To detect path of directional preference (DP), active single and repeated spinal movements in standing, and sustained positions in standing and prone position were assessed [23]. Passive hip internal rotation range of motion (ROM) was measured using a goniometer in prone position [34]. Spinal mobility and pain arising from applying spring test from L1 to L5 spinous processes was recorded. If any pain was produced by spring test, the prone instability test (PIT) was done afterward [35].

Traditional classification based on 2007 TBC algorithm is consisted of 2 stages. Initially, 8 criteria sets hierarchically to guide the therapist assigning the patients to specific exercise, manipulation and stabilization subgroups, respectively. If a patient did not meet criteria of any subgroups in stage one, in the second stage 13 additional criteria is conveyed. In this stage, the therapist determined the patient's best fitted subgroup based on analyzing factors favoring and against each subgroup's criteria. All TBC criteria are listed in Table 1.

Statistical analysis: To assess the strength of 2007 TBC system, an equivalent clinical C3 CA typology was conducted using PAM analysis. Because the kappa coefficient is a chance-corrected measure of agreement, it is the appropriate reliability coefficient for evaluating the reliability of the 2007 TBC system. Therefore, Cohen's kappa was run to examine the level of agreement between the TBC system and the C3 cluster typology [36]. The Silhouette width was used to assess the strength of clustering. The Silhouette width measures a combination of intra cluster homogeneity and inter-cluster heterogeneity. Discriminative power and prevalence of selected criteria was determined for each variable in each cluster. By means of a one-sided chai-square test it was determined whether the frequency of positive variables differed significantly per cluster. All statistical analyses were performed in R 3.5.0 and SPSS 19.0 software. Also for PAM analysis VarSelLCM packages of R software (R Core Team, Ver. 3.5.3, New Zealand) were used.

Results

Distribution of quantitative variables within TBC subgroups presented in Table 2 for a total sample of 90 participants (50 males and 40 females). The ratios of male/female in TBC system were 24/15 for DP flexion, 16/10 for manipulation and 10/15 for stabilization subgroups.

PAM analysis revealed the evidence of clustering for a C3 cluster typology (Average Silhouette width 0.12). Discriminative power and prevalence of selected criteria for C3 typology was shown in Table 3. The prevalence

Table 1: All 2007 Treatment Based Classification (TBC) criteria included for Partitioning Around Medoids (PAM) analysis.

8 primary criteria of the TBC system	13 additional criteria of the TBC system
• Age<40	• Increasing episode frequency
• Symptom duration<16d	• Segmental hypermobility in spring test
• Centralization with flexion and peripheralization with extension	• Segmental hypomobility in spring test
• Centralize with 2 or more movements in the same direction (i.e., flexion or extension)	• 3 or more episodes
• Pain distribution	• Low FABQ scores (FABQ-W <19)
• PIT positive	• Hip medial rotation ROM >35º
• SLR ROM>91°	• Peripheralization with motion testing
• Aberrant motion	• No pain with spring test
	• Discrepancy in SLR ROM (>10º)
	• (FABQ-PA <9)
	• Low back pain only (no distal symptoms)
	• Status quo with all Movements
	• Directional preference

TBC: Treatment Based Classification, FABQ: Fear Avoidance Belief Questionnaire, PA: Physical Activity Subscale, PIT: Prone Instability Test, SLR ROM: Straight Leg Raising Range of Motion

Table 2: Quantitative variables within subgroups of the treatment based classification (TBC) system.

Variable	Total sample (n=90)	2007 TBC subgroups		
		DP Flex (n=39)	Manipulation (n=26)	Stabilization (n=25)
Age	45.78 ± 13.13	50.51 ± 11.43	45.58 ± 14.37	38.6 ± 11.28
BMI	26.27 ± 4.15	27.33 ± 4.09	24.42 ± 3.8	26.53 ± 4.11
Pain intensity	5.83 ± 1.45	5.99 ± 1.33	4.92 ± 1.41	6.54 ± 1.23
ODQ score (%)	38.22 ± 15.43	40.08 ± 14.98	33.90 ± 17.23	40.67 ± 12.87
FABQ Total	58.57 ± 17.77	62.95 ± 16.11	50.61 ± 17.8	60 ± 18.13
FABQ-W	25.17 ± 9.04	26.46 ± 8.06	22.69 ± 9.48	25.72 ± 9.83
FABQ-PA	19.71 ± 4.32	20.59 ± 3.85	17.58 ± 5.01	20.56 ± 3.54

TBC: Treatment Based Classification, DP Flex: Direction of Preference in Flexion, BMI: Body Mass Index, ODQ: Oswestry Disability Questionnaire, FABQ: Fear Avoidance Belief Questionnaire, PA: Physical Activity Subscale

Table 3: Discriminative power and prevalence of selected criteria by Partitioning around Medoids (PAM) analysis for C3 cluster typology.

PAM Selected criteria	Discriminative power (%)	Cumulative discriminative power (%)	Prevalence in each cluster (%)		
			Cluster 1 (n=46) (%)	Cluster 2 (n=7) (%)	Cluster 3 (n=37) (%)
TBC subgroups prevalence	-	-	DP Flex 63 Manipulation 26 Stabilization 11	DP Flex 14 Manipulation 71 Stabilization 14	DP Flex 24 Manipulation 24 Stabilization 51
Age<40	29.64	29.64	No 100	Yes 57	Yes 84
Increasing episode frequency	15.04	44.68	Yes 100	No 86	Yes 100
Symptom duration<16d	10.91	55.59	No 100	Yes 71	No 97
Centralization with flexion and peripheralization with extension	10.28	65.88	Yes 65	No 100	No 78
Pain distribution	7.83	73.71	Below knee 56	Above knee 100	Above knee 78
PIT positive	7.74	81.45	No 67	No 57	Yes 68
Segmental hypermobility	5.38	86.83	No 74	No 86	Yes 57
3 or more episodes	5.31	92.14	Yes 76	No 71	Yes 73
SLR ROM>91°	4.01	96.15	No 61	No 100	No 61
Aberrant motion	3.85	100	No 78	No 86	No 62

PAM: Partitioning Around Medoids, TBC: Treatment Based Classification, DP Flex: Direction of Preference in Flexion, PIT: Prone Instability Test, SLR ROM: Straight Leg Raising Range of Motion.

of the TBC subgroups within C3 clusters was also reported. The majority of patients in the cluster 1 (63%) were those with DP flexion based on the TBC system. Also, the majority of patients in cluster 2 (71%) were the manipulation subgroup and about 51% of cluster 3 were the stabilization exercise subgroup. The highest heterogeneity was observed in cluster 3 (Table 3).

The estimation of correlation between each variable and individual subgroups were further evaluated by considering overlap between the ranges of variable's result (from 0-100%) for each subgroup in adjoining clusters (Table 3). Accordingly, the lower the overlap, the higher the correlation, i.e. the amount of 0% or 100% for the results of variable of "age<40", indicates the minimum amount of overlap and therefore the maximum amount of correlation. The maximum amount of overlap with minimum correlation is where the results of the variable were achieved to 50%.When the patients were divided into 3 clusters, these overlaps were considerable and the average of Silhouette width is low. In the C3 typology, 10 out of 21 TBC criteria were selected by PAM that could differentiate 100% of patients. None of the criteria associated with fear-avoidance beliefs were included in the new C3 typology.

Cohen's Kappa revealed fair agreement be-

tween the TBC system and C3 cluster typology (Percent of agreement 61%, Kappa=0.36, p <0.001). The results of chai-square test showed that despite the fact that there was no significant difference in the prevalence of "SLR ROM" and "aberrant motion" between C3 clusters, these criteria were selected by PAM as effective criteria for discrimination (Table 3). Therefore, in the C3 typology, 5 symptom criteria, 3 mobility and 2 control criteria were included to differentiate between clusters.

Discussion

The original 2007 TBC criteria by PAM analysis was employed to evaluate the agreement of original TBC and new CA system. Our results revealed that the TBC system has fair agreement with the new 3 cluster typology with 61% of agreement. The lower values of kappa are attributable to increase the probability of chance agreement between two classification approaches [36].

In this study, selected criteria by PAM analysis were different with original TBC system. Main subgrouping criteria for original TBC system were the primary 8 criteria, setting hierarchically to assign patients to subgroups. Main PAM selected criteria for C3 typology also were those of 7 primary criteria of TBC system with higher discriminative power. This indicates the logical agreement between the TBC system and the C3 clustering. However, the percent of agreement in C3 typology is not a desirable limit for a clinical clustering model [37]. If we consider the PAM analysis as a valid classification method, this relatively low agreement revealed the weakness of the 2007 TBC system.

A moderate [23-26] and moderate to good [21, 29, 38, 39] inter-observer reliability have been reported for the TBC system. To the authors' knowledge, this is the first study that examine the TBC classification using clustering by PAM analysis. All previous studies evaluated the inter-observer or the intra-observer reliability of the TBC approach. Therefore, their results cannot be compared with our results.

Stanton et al. showed that the TBC criteria do not guarantee that patients comply with a single intervention, but that 25% of patients can meet criteria for more than one subgroup [23]. In addition, the prevalence of patients with "unclear classification" in early versions of TBC system has been estimated between 26 and 50 percent [23, 24, 38, 40]. The developers of the 2007 TBC system also stated that this system had its limitations and needed to refine the algorithm [22]. In line with previous studies [23, 40, 41], our results disclosed the evidence of within group inhomogeneity and between group overlapping in the 2007 TBC system. Decision making based on C3 system is inconclusive and it is not comprehensive enough to consider the various clinical presentations of patients with LBP. This may be due to lower average silhouette width of 0.12 and omission of more informative clinical criteria such as aberrant motion, positive PIT test and SLR ROM. The PAM analysis also omitted the results of the spring test in the clustering. These criteria provided beneficial information about the mobility of the lumbar spine. The imperfect combination of criteria in the C3 typology uncover the multifaceted interaction of the variables. It seems that the patients with LBP could be classified into more than 3 clusters to address all presentations of LBP.

The results of Chai-square test showed that the relying only on the traditional statistical methods and their p-values cannot determine whether a criterion is effective or not in a classification. These results are in agreement with recent statistical paradigms related to concept of "statistical significance" [42, 43]. According to recent paradigms, we cannot interpret the results based on p-values only. Thus, the phrase of "no significant" does not mean that there is "no relationship" or "no effectiveness" for an intervention approach. We need other statistical indices for better analysis and interpretation of the results [42, 43].

Since, all variables in the TBC system were examined individually and separately, the behavior of one variable in the presence or absence of the others, may affect the final outcome of the classification. The CA approach in our study may be contrasted against traditional multivariate methods [21, 23-26, 38-41] that typically analyze for significant differences between variables. In many applied research studies, the most useful level of analysis involves the consideration of the entire case more than comparisons between variables within each case [1]. Clustering tries to discover and identify unknown patterns without any limitation. PAM analysis is one advantages of this study that could handle the mixed data sets without challenging [1, 4].

One limitation of the present study was the inclusion of 21 criteria in the clustering for the sample size of 90 patients that may affect the strength of C3 typology. Another limitation was related to inherent weakness of the 2007 TBC system. Due to the limited examination criteria and the lack of reliable tests in the diagnosis of movement control disorders in the 2007 TBC system, the C3 typology failed to identify properly the criteria of the stabilization exercise subgroup.

Conclusion

Higher probability of chance agreement was observed between two classification methods. The classification of patients with LBP into C3 typology decreased the strength of clustering markedly. The C3 cluster typology provides the evidence that LBP is a heterogeneous condition with vast combination of signs and symptoms. To that end, traditional classification of patients without considering simultaneous interaction of selected criteria may lead to erroneous results.

Acknowledgment

This article has been derived from the research done by Esmaeil Shokri as a part of the PhD thesis and financially supported by Shiraz University of Medical Sciences, Shiraz, Iran (Grant No 12195-06-01-95).

Conflict of Interest

None

References

1. Leonard ST, Droege M. The uses and benefits of cluster analysis in pharmacy research. *Res Social Adm Pharm*. 2008;**4**(1):1-11. doi: 10.1016/j.sapharm.2007.02.001. PubMed PMID: 18342818.
2. Davidoff F. Heterogeneity Is Not Always Noise: Lessons From Improvement. *JAMA*. 2009;**302**(23):2580-6. doi: 10.1001/jama.2009.1845.
3. Armstrong JJ, Zhu M, Hirdes JP, Stolee P. K-means cluster analysis of rehabilitation service users in the Home Health Care System of Ontario: examining the heterogeneity of a complex geriatric population. *Arch Phys Med Rehabil*. 2012;**93**(12):2198-205. doi: 10.1016/j.apmr.2012.05.026. PubMed PMID: 22705468.
4. Madhulatha TS. An overview on clustering methods. *arXiv preprint arXiv:12051117*. 2012.
5. Fourney DR, Andersson G, Arnold PM, Dettori J, Cahana A, Fehlings MG, et al. Chronic low back pain: a heterogeneous condition with challenges for an evidence-based approach. *Spine (Phila Pa 1976)*. 2011;**36**(21 Suppl):S1-9. doi: 10.1097/BRS.0b013e31822f0a0d. PubMed PMID: 21952181.
6. Furlan AD, Tomlinson G, Jadad AA, Bombardier C. Examining heterogeneity in meta-analysis: comparing results of randomized trials and nonrandomized studies of interventions for low back pain. *Spine (Phila Pa 1976)*. 2008;**33**(3):339-48. doi: 10.1097/BRS.0b013e31816233b5. PubMed PMID: 18303468.
7. McKenzie R, May S. The Lumbar Spine Mechanical Diagnosis & Therapy. 2nd ed. Orthopedic Physical Therapy Products; 2003.
8. Delitto A, Erhard RE, Bowling RW. A treatment-based classification approach to low back syndrome: identifying and staging patients for conservative treatment. *Phys Ther*. 1995;**75**(6):470-85. doi: 10.1093/ptj/75.6.470. PubMed PMID: 7770494.
9. Sahrmann S, Azevedo DC, Dillen LV. Diagnosis and treatment of movement system impairment

syndromes. *Braz J Phys Ther.* 2017;**21**(6):391-9. doi: 10.1016/j.bjpt.2017.08.001. PubMed PMID: 29097026. PMCID: PMC5693453.

10. O'Sullivan P. Diagnosis and classification of chronic low back pain disorders: maladaptive movement and motor control impairments as underlying mechanism. *Man Ther.* 2005;**10**(4):242-55. doi: 10.1016/j.math.2005.07.001. PubMed PMID: 16154380.
11. Maluf KS, Sahrmann SA, Van Dillen LR. Use of a classification system to guide nonsurgical management of a patient with chronic low back pain. *Phys Ther.* 2000;**80**(11):1097-111. PubMed PMID: 11046197.
12. O'Sullivan PB. Lumbar segmental 'instability': clinical presentation and specific stabilizing exercise management. *Man Ther.* 2000;**5**(1):2-12. doi: 10.1054/math.1999.0213. PubMed PMID: 10688954.
13. Hefford C. McKenzie classification of mechanical spinal pain: profile of syndromes and directions of preference. *Man Ther.* 2008;**13**(1):75-81. doi: 10.1016/j.math.2006.08.005. PubMed PMID: 17188547.
14. Hirsch O, Strauch K, Held H, Redaelli M, Chenot JF, Leonhardt C, et al. Low back pain patient subgroups in primary care: pain characteristics, psychosocial determinants, and health care utilization. *Clin J Pain.* 2014;**30**(12):1023-32. doi: 10.1097/AJP.0000000000000080. PubMed PMID: 24480909.
15. Viniol A, Jegan N, Hirsch O, Leonhardt C, Brugger M, Strauch K, et al. Chronic low back pain patient groups in primary care--a cross sectional cluster analysis. *BMC Musculoskelet Disord.* 2013;**14**:294. doi: 10.1186/1471-2474-14-294. PubMed PMID: 24131707; PMCID: PMC3852748.
16. Billis E, McCarthy CJ, Roberts C, Gliatis J, Papandreou M, Gioftsos G, et al. Sub-grouping patients with non-specific low back pain based on cluster analysis of discriminatory clinical items. *J Rehabil Med.* 2013;**45**(2):177-85. doi: 10.2340/16501977-1100. PubMed PMID: 23321847.
17. Reme SE, Shaw WS, Steenstra IA, Woiszwillo MJ, Pransky G, Linton SJ. Distressed, immobilized, or lacking employer support? A subclassification of acute work-related low back pain. *J Occup Rehabil.* 2012;**22**(4):541-52. doi: 10.1007/s10926-012-9370-4. PubMed PMID: 22644216.
18. Karayannis NV, Jull GA, Hodges PW. Physiotherapy movement based classification approaches to low back pain: comparison of subgroups through review and developer/expert survey. *BMC Musculoskelet Disord.* 2012;**13**:24. doi: 10.1186/1471-2474-13-24. PubMed PMID: 22348236; PMCID: PMC3395852.
19. Brennan GP, Fritz JM, Hunter SJ, Thackeray A, Delitto A, Erhard RE. Identifying subgroups of patients with acute/subacute "nonspecific" low back pain: results of a randomized clinical trial. *Spine (Phila Pa 1976).* 2006;**31**(6):623-31. doi: 10.1097/01.brs.0000202807.72292.a8. PubMed PMID: 16540864.
20. Fritz JM, Delitto A, Erhard RE. Comparison of classification-based physical therapy with therapy based on clinical practice guidelines for patients with acute low back pain: a randomized clinical trial. *Spine (Phila Pa 1976).* 2003;**28**(13):1363-71. doi: 10.1097/01.brs.0000067115.61673.ff. PubMed PMID: 12838091.
21. Fritz JM, Cleland JA, Childs JD. Subgrouping patients with low back pain: evolution of a classification approach to physical therapy. *J Orthop Sports Phys Ther.* 2007;**37**(6):290-302. doi: 10.2519/jospt.2007.2498. PubMed PMID: 17612355.
22. Alrwaily M, Timko M, Schneider M, Stevans J, Bise C, Hariharan K, et al. Treatment-Based Classification System for Low Back Pain: Revision and Update. *Phys Ther.* 2016;**96**(7):1057-66. doi: 10.2522/ptj.20150345. PubMed PMID: 26637653.
23. Stanton TR, Fritz JM, Hancock MJ, Latimer J, Maher CG, Wand BM, et al. Evaluation of a treatment-based classification algorithm for low back pain: a cross-sectional study. *Phys Ther.* 2011;**91**(4):496-509. doi: 10.2522/ptj.20100272. PubMed PMID: 21330450.
24. Apeldoorn AT, van Helvoirt H, Ostelo RW, Meihuizen H, Kamper SJ, van Tulder MW, et al. Inter-rater reliability of a modified version of Delitto et al.'s classification-based system for low back pain: a pilot study. *J Man Manip Ther.* 2016;**24**(2):98-110. doi: 10.1179/2042618614Y.0000000082. PubMed PMID: 27559279. PMCID: PMC4984810.
25. Alyazedi FM, Lohman EB, Wesley Swen R, Bahjri K. The inter-rater reliability of clini-

cal tests that best predict the subclassification of lumbar segmental instability: structural, functional and combined instability. *J Man Manip Ther.* 2015;**23**(4):197-204. doi: 10.1179/2042618615y.0000000002. PubMed PMID: 26917937. PMCID: PMC4727732.

26. Fritz JM, George S. The use of a classification approach to identify subgroups of patients with acute low back pain. Interrater reliability and short-term treatment outcomes. *Spine (Phila Pa 1976).* 2000;**25**(1):106-14. doi: 10.1097/00007632-200001010-00018. PubMed PMID: 10647168.
27. Petersen T, Olsen S, Laslett M, Thorsen H, Manniche C, Ekdahl C, et al. Inter-tester reliability of a new diagnostic classification system for patients with non-specific low back pain. *Aust J Physiother.* 2004;**50**(2):85-94. doi: 10.1016/s0004-9514(14)60100-8. PubMed PMID: 15151492.
28. Fritz JM, Irrgang JJ. A comparison of a modified Oswestry Low Back Pain Disability Questionnaire and the Quebec Back Pain Disability Scale. *Phys Ther.* 2001;**81**(2):776-88. doi: 10.1093/ptj/81.2.776. PubMed PMID: 11175676.
29. Beurskens A, De Vet HCW, Köke AJA. Responsiveness of functional status in low back pain: a comparison of different instruments. *Pain.* 1996;**65**(1):71-6. doi: 10.1016/0304-3959(95)00149-2. PubMed PMID: 8826492.
30. Mousavi SJ, Parnianpour M, Mehdian H, Montazeri A, Mobini B. The Oswestry Disability Index, the Roland-Morris Disability Questionnaire, and the Quebec Back Pain Disability Scale: translation and validation studies of the Iranian versions. *Spine (Phila Pa 1976).* 2006;**31**(14):E454-9. doi: 10.1097/01.brs.0000222141.61424.f7. PubMed PMID: 16778675.
31. Jensen MP, Turner JA, Romano JM. What is the maximum number of levels needed in pain intensity measurement? *Pain.* 1994;**58**(3):387-92. doi: 10.1016/0304-3959(94)90133-3. PubMed PMID: 7838588.
32. Rostami M, Noorian N, Mansournia MA, Sharafi E, Babaki AE, Kordi R. Validation of the Persian version of the fear avoidance belief questionnaire in patients with low back pain. *J Back Musculoskelet Rehabil.* 2014;**27**(2):213-21. doi: 10.3233/bmr-130439. PubMed PMID: 24254494.
33. Waddell G, Somerville D, Henderson I, Newton M. Objective clinical evaluation of physical impairment in chronic low back pain. *Spine (Phila Pa 1976).* 1992;**17**(6):617-28. doi: 10.1097/00007632-199206000-00001. PubMed PMID: 1308095.
34. Nussbaumer S, Leunig M, Glatthorn JF, Stauffacher S, Gerber H, Maffiuletti NA. Validity and test-retest reliability of manual goniometers for measuring passive hip range of motion in femoroacetabular impingement patients. *BMC Musculoskelet Disord.* 2010;**11**:194. doi: 10.1186/1471-2474-11-194. PubMed PMID: 20807405. PMCID: PMC2942800.
35. Hicks GE, Fritz JM, Delitto A, Mishock J. Interrater reliability of clinical examination measures for identification of lumbar segmental instability. *Arch Phys Med Rehabil.* 2003;**84**(12):1858-64. doi: 10.1016/s0003-9993(03)00365-4. PubMed PMID: 14669195.
36. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas.* 1960;**20**(1):37-46. doi: 10.1177/001316446002000104.
37. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med.* 2005;**37**(5):360-3. PubMed PMID: 15883903.
38. Henry SM, Fritz JM, Trombley AR, Bunn JY. Reliability of a treatment-based classification system for subgrouping people with low back pain. J Orthop Sports *Phys Ther*. 2012;**42**(9):797-805. doi: 10.2519/jospt.2012.4078. PubMed PMID: 22677525.
39. Fritz JM, Brennan GP, Clifford SN, Hunter SJ, Thackeray A. An examination of the reliability of a classification algorithm for subgrouping patients with low back pain. *Spine (Phila Pa 1976).* 2006;**31**(1):77-82. doi: 10.1097/01.brs.0000193898.14803.8a. PubMed PMID: 16395181.
40. Apeldoorn AT, Ostelo RW, van Helvoirt H, Fritz JM, Knol DL, Van Tulder MW, et al. A randomized controlled trial on the effectiveness of a classification-based system for subacute and chronic low back pain. *Spine (Phila Pa 1976).* 2012;**37**(16):1347-56. doi: 10.1097/BRS.0b013e31824d9f2b. PubMed PMID: 22333955.
41. Stanton TR, Hancock MJ, Apeldoorn AT, Wand BM, Fritz JM. What characterizes people who have an unclear classification using a treatment-based classification algorithm

for low back pain? A cross-sectional study. *Phys Ther.* 2013;**93**(3):345-55. doi: 10.2522/ptj.20120263. PubMed PMID: 23139426. PMCID: PMC3588107.

42. Wasserstein RL, Schirm AL, Lazar NA. Moving to a World Beyond "p<0.05". *The American Statistician.* 2019;**73**(supp1):1-19. doi: 10.1080/00031305.2019.1583913.

43. Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. *Nature*. 2019;**567**(7748):305-7. doi: 10.1038/d41586-019-00857-9. PubMed PMID: 30894741.