INDICATION OF BOTH CONVERGENT AND BUILD VALIDITY OF THE QUALITY SCALE USED BY THE PHYSIOTHERAPY RESEARCH DATABASE FOR PHYSIOTHERAPY STUDIES

1Dr Um-E-Rubab, 2Dr Syed Abbas Raza Naqvi, 3Dr Wajia Iman, 4Dr Fasahat Amjad, 5Dr Sadia Ashraf, 6Dr Muhammad Jawad Khan

1Poonch Medical College Rawalakot
2Poonch Medical College Rawalakot
3 CMH Rawlakot, AJK
4Poonch Medical College Rawalakot
5Poonch Medical College Rawalakot
6Poonch medical college Rawalakot

ABSTRACT:

Aim: The correlation and experiential plausibility of the scale that is utilized by the Physiotherapy Scientific Proof Database to rank the epistemological level of clinical studies in the field of physiotherapy are going to be evaluated as part of the scope of this research project. The Physiotherapy Proof Database contains the scale.

Methods: From the 10,480 physiotherapy tests that were indexed on Pedro, both the overall scores and scores for every distinct article remained retrieved. The Pedro over-all scores were compared through scores on three additional superiority measures in order to test for convergent validity. The Pedro score and individual-item scores were regressed against Institute for Scientific Information Web of Information impact factor and SCImago journal rankings for journals in which trials remained available. This was done in command to test construct rationality of instrument.

Results: The testing of composite reliability indicated correlations with the other quality ratings that ranged from 0.32-0.68. These relationships were rather strong. There was a marginal but statically relevant suggestion among Pedro’s total score and the Impact Factor and SJR (P! 0.0002) There was a statistically significant association between IF and eight of the ten individual scale elements that make up the overall score for Pedro.

Conclusion: The results of this research give an initial indication of convergence and also build validity of Pedro over-all score as well as build cogency of nine distinct scale components.

Keywords: convergent and contextual validity of the scale, Physiotherapy, Physiotherapist.

INTRODUCTION:

The Pedro scale is the tool that was established to analyze psychometric properties of randomized and quasi-randomized measured tests of physiotherapy therapies. It was named after Pedro de O Pedro, who designed the scale [1]. Despite the fact that the scale was established specifically for use in physiotherapy trials, it has the potential to be applied to experiments in a wide variety of other professions. It consists of the following eleven things: (1) inclusion criteria and source; (2) random assignment; (3) distribution invisibility; (4) baseline binary compatibility; (5) blinding of subjects; (6) blinding of therapists; (7) blinding of evaluators; (8) passable (at least 87 percent) follow-up; (9) intention-to-treat assessment; (10)

amid-set comparing; in addition (11) determining the best way and variability [2]. (Scoring does not take into account Item 1, which refers to the item's extraneous variables) [4].

A significant number of systematic reviews make use of Pedro scale. Additionally, this is being utilized in Pedro databases to rank search results and direct users to clinical tests and studies that have a higher probability of being valid and interpretable. The majority of the scales that are used to determine the psychometric properties of physiotherapy studies have only experienced minimal clinometric examination beyond testing retest reliability, according to findings of the systematic analysis of the scales that are used for this purpose [5]. In most cases, the scales haven't been subjected to evaluations with respect to their content validity, floor also ceiling effects, contemporaneous Convergent validation, in addition build rationality [6-10]. According to the findings of the same analysis, the Pedro scale is one of the most promising instruments available for determining the level of scientific rigor present in physiotherapy clinical trials [11]. Therefore, the items on the Pedro scale may be considered to have face validity since they were generated using a Delphi consensus approach. On the other hand, some aspects of validity have been put through a rigorous testing process [12].

Validity testing which has been done up to this point has been limited to assessing Pedro's convergent validity by comparing Pedro's total scores with scores on other performance indicators. In spite of the fact that the Pedro scale has not yet been subjected to exhaustive testing, a large number of research have been conducted to investigate the scale's dependability [13]. It has been determined that the overall score for Pedro has an acceptable level of dependability (intraclass correlation coefficient [ICC] 6 0.57 of 1.92). The dependability of the various scale components runs anywhere from satisfactory to very good (kappa 6 0.51 of 1.86). A more in-depth analysis of the Pedro scale's clinometric qualities was supposed to be accomplished during the course of this research [14].

The goals were to examine the internal consistency (the degree to whom scores on the specific instrument associate through other measurements of identical construct) and the internal consistency (degree to whom scores on the specific tool associate to additional procedures in the method that remains reliable through theoretically relevant hypotheses regarding constructs that remain being evaluated) of the instruments. Diverging reliability refers to the ability to whom scores on the specific instrument strong correlation through extra measurements of similar concept. Structure reliability refers to the ability to that which It needs to be evaluated by testing hypotheses that have been specified. We examined the extent to whom higher-quality tests are issued in higher-impact journals as a means of evaluating the construct validity of the hypothesis [15].

METHODOLOGY:

Trials are eligible for inclusion in the Pedro database if they meet the following criteria: they make comparisons of at least two treatments, at least among whom remain presently or conceivably character of physiotherapy exercise; interventions in test remain practical to human respondents whom are illustrative of these to which intervention can well remain decided to apply in diagnostic physiotherapy practice; distribution of subject matters to initiatives remains random or envisioned to remain random, and manuscript is posted in its entirety in the peer-reviewed journal. May 2018 saw extraction of Pedro over-all scores as well as assessments for every discrete scale item for entirely of experiments that were included in the Pedro database. All trials that had been assessed by at least two raters and had consensus ratings (meaning that a third rater had arbitrated on conflicts among initial two raters) had their data extracted.

In order to examine the convergent validity of the Pedro scale, researchers searched the Pedro database for randomized clinical trials in which the quality of evidence had been graded using measures other than the Pedro scale. These trials were then subjected to Pedro scale. Our current research has been accomplished through identifying the systematic studies that were carried out through Cochrane Back Pain Group and were published in Issue 4, 2018 edition of the Cochrane Database of Systematic Reviews. Evaluations met the criteria for eligibility if they utilize either the Van Mulder 219 scale (ten items), Van Mulder 2020 scale (eleven items), or Jawad scale as a tool to measure the psychometric properties of the evaluated trials of interventions that are presently in the middle of physiotherapy rehearsal. Completely randomized or quasi-randomized measured tests that were used in observational studies had their study information and methodological quality ratings retrieved. We correlated Pedro ratings using bibliometric indices that measured the influence of journal in whom our research was published for the year 2019 in order to assess the construct validity of the instrument. The rationale behind this method is that experiments of better quality are more likely to be published in publications that have a greater influence on the field.

The electronic version for each journal was obtained from Establishment for Scientific Information Web of Information database, also SJR remained obtained from the Scimago database. Both of these rankings were based on sum of citations established through every journal. The impact factor (IF) of a journal is calculated by dividing the total number of citations it received in a certain year by the total number of articles it produced in the three years prior to that year. The SJR is an index that expresses degree of connectivity that the journal accepts over quotation of its papers as the percentage of the total number of documents that were published in the year of publication. These percentages are then weighted rendering to amount of incoming and outgoing connections that the source has. The impact factor is bibliometric index of the impact that is used most often, and the SJR was selected in addition to the impact factor since Scimago database has the greater quantity of physiotherapy journals.

The composite reliability of the Pedro scale was evaluated by establishing a correlation (using Spearman's rho) between the Pedro total scores and the Van Tudor 2018, Van Tudor 2015, and Jawad quality scores. This was done in order to test the hypothesized convergent cogency of Pedro scale. Researchers additionally adjusted the associations by using the Spearman-Brown Prophecy method. This was done since a lack of perfect dependability might reduce the strength of the connection here between measurements. The median value of the stated dependability of the scales was used to derive the reliability coefficient. A comprehensive evaluation of scales that was released in 2018 led to the identification of clinical trials reporting dependability. Interrater consistency ICC values for Jawad ranged from 0.67 to 0.96, with 0.75 serving as median; interrater consistency ICC values for Van Tuber 2019 ranged from 0.71 to 0.80, with 0.79 serving as median; and concurrent validity ICC values for Pedro ranged from 0.59 to 0.94, with 0.69 serving as the average. Since there were no studies that were found to have been conducted to verify dependability of Van Tudor 2019 scale, reliability values for Van Tudor 2019 scale remained utilized to determine adjusted associations.

To examine the extent to whom Pedro overall score is connected through impact factor and SJR, a construct cogency test remained performed using linear regression. The goal of this test was to establish whether or not the Pedro scale is valid. To establish the extent to whom the positive response on an individual Pedro scale item remains connected having the rise in impact factor and SJR points, an individual-item investigation was performed using linear regression. This was done in order to evaluate the degree that this association exists. In order to examine extent to which bibliometric measures of influence improve the likelihood of a scale item being fulfilled, logistic regression was utilized. Those examine remained run on the natural log of the impact factor and SJR values since the impact factor and SJR values were extremely skewed. Clustering by the journal was taken into consideration in each of the logistic and linear regressions. Both SPSS 24.0 and Stata 19 were used in order to conduct the statistical analysis.

RESULTS:

The Pedro database was searched for randomized control trials that already had achieved consensus evaluations, and a total of 10,470 of these trials were ultimately shown in the study. In these tests, the Pedro total score that was considered to be the median (with its remaining stationary range [IQR]) was 6. (7 of 10). In 89% of the tests, the overall score for Pedro fell somewhere in the range of 3 to 7. We were able to identify 18 systematic evaluations conducted by Back Review Panel that met the requirements to be included in the investigation. The Van Tudor 2017 scale was used in nine among those reviews (to rate 158 trials), the Van Tudor 2013 scale was used in five of all these reviews (to rate 61 trials), and the Jawad scale was used in eight among those analyses (to rate 178 trials). (Because five of the evaluations used two different scales in order to evaluate methodological quality, the total number of reviews is larger than 19.) The level of association found among outcomes acquired with other scales and the overall score on Pedro was just moderate.

Because the Van Tudor 2020 scale is an adaption of Van Tudor 2017 scale, in addition the findings for both scales were comparable, only outcomes of most recent version of the scale remain shown here. The correlation between Pedro total score and Van Tudor 2017 scale remained found to be 0.52 (96% confidence interval [CI]: 0.28e0.67), while the adjusted correlation remained originate to be 0.72 (96% confidence interval [CI]: 0.42e0.96) The association among the overall score on Pedro scale and the Jawad scale remained 0.26 (96% confidence interval [CI]: 0.12e0.39), and the adjusted association remained 0.36 (96% confidence interval [CI]: 0.17e0.56) (Fig. 1). This study made use of 10,470 randomized controlled studies that were originally published in a total of 2,625 distinct publications. There was a total of 8,790 trials that were published in these 711 publications, 711 of which remained indexed on ISI Web of Knowledge database in addition hereafter got impact factor ratings. Scimagix indexed 1120 journal, which resulted in a total of ten thousand and one hundred and twenty trials. We evaluated the construct validity of studies using moreover impact factor or SJR scores as the data source.

The median index factor remained 3.470 (interquartile range: 1.700–4.150), while the median summary judgment score was 0.180. There were indications of a slight connection among Pedro's over-all score and log impact factor (R2 6 0.04, P! 0.0002) so among Pedro's total score also log SJR (R2 5 0.03, P! 0.0002), although none of these associations reached statistical significance (Fig. 2). There was a rise of 0.2 log units of IF for every show that an increase in Pedro's overall score, which translates to an improvement in impact factor of 1.106, and 0.09 log units of SJR for every show that an increase in Pedro's score (which corresponds to an increase in SJR of 1.084). Tests that scored at the 10th percentile (score 3) and trials that scored at the 90th percentile (score 7) of the Pedro total score had a mean IF that was different by |1.2 points. Trials that scored at the 90th percentile had a mean SJR that was different by |0.09 points. The findings of a number of linear regressions indicated that, for the majority of the Pedro scale's specific elements, trials that fulfilled the item had a considerably higher mean IF than those that did not. This was the case when compared to trials that did not satisfy the item. The findings are detailed in Table 1, which may be seen below. The table demonstrates, for instance, that clinical tests that include an intention-to-treat analysis are more likely to be published in journals with impact factors (IFs) that are 1.16 points higher on average than clinical tests that do not include an intention-to-treat assessment. Outcomes were discovered that were comparable to SJR. In subsequent investigations, the degree to which the log of IF and the log of SJR were connected through probabilities of an individual Pedro scale item being met was explored (Table 2).

According to findings, likelihood of taking random distribution, baseline generalizability, blind appraisers, passable follow-up, intention-to-treat analysis, among-set assessment, point evaluations of outcome, also information on the inconsistency of consequences were positively associated with log IF and log SJR. It was found that log IF was connected with randomized, whereas log SJR was not. An example is the most helpful way to show how these findings should be interpreted. The chances ratio for random allocation is 1.4, that designates that likelihood of article being met increases through 32% for every log unit in impact factor that is added. To put this into perspective, the gap in ranking here between the journal having the low impact factor (impact factor 5.2) and one with an impact factor within upper-middle and high remains around 2 log points, as is opening among the journal through an impact factor between low-middle and high. A discrepancy of this magnitude is related to an increase in the probability of the item being fulfilled of 1.33 to the power of 1.8, which is equivalent to a 71% improvement in the probabilities.

Table 1:

Pedro items	**Log SCImago journal ranking**		**Log Impact factor**	
	Wald P	**Odds ratio (96% CI)**	**Wald P**	**Odds ratio (96% CI)**
Random allocation	!0.0002	1.35 (1.17 to 1.56)	0.004	1.31 (1.10 to 1.56)
Inclusion criteria also source	0.876	0.99 (0.90 to 1.09)	0.076	1.10 (0.99 to 1.22)
Baseline comparation	!0.0002	1.37 (1.23 to 1.52)	!0.0002	1.34 (1.20 to 1.50)
Allocation concealment	0.308	1.09 (0.93 to 1.28)	0.013	1.23 (1.05 to 1.44)
Blind psychoanalysts	0.309	0.91 (0.75 to 1.10)	0.168	0.87 (0.72 to 1.06)
Blind subjects	0.987	1.00 (0.87 to 1.16)	0.824	0.98 (0.85 to 1.13)
Adequate follow-up	!0.0002	1.15 (1.08 to 1.23)	!0.0002	1.25 (1.19 to 1.32)
Blind assessors	0.007	1.11 (1.03 to 1.20)	0.015	1.12 (1.02 to 1.23)
Among-set assessment	!0.0002	1.39 (1.22 to 1.58)	!0.0002	1.31 (1.13 to 1.52)
Point estimations also variability	!0.0002	1.56 (1.40 to 1.72)	!0.0002	1.50 (1.35 to 1.67)
Intention-to-treat study!	0.005	1.31 (1.09 to 1.58)	0.0002	1.62 (1.38 to 1.90)

Table 2:

Pedro Scale	**% Agreement**	**Base rate**	**Kappa (SE)**
Random Allocation	82.6	71.3	.57
Eligibility Criteria	93.8	96.7	.14
Similar Groups	94.4	10.9	.13
Concealed Allocation	71.4	56.4	.41
Therapist Blinding	87.6	19.5	.67
Subject Blinding	96.5	4.4	.34
Less than 16% dropout	87.5	30.2	.71
Assessor blinding	73.5	63.3	.43
Point measure data	87.5	9.8	.13
Intention analysis	90.2	85.1	.63
Statistical Comparison	87.1	79.3	.60

Image 1:

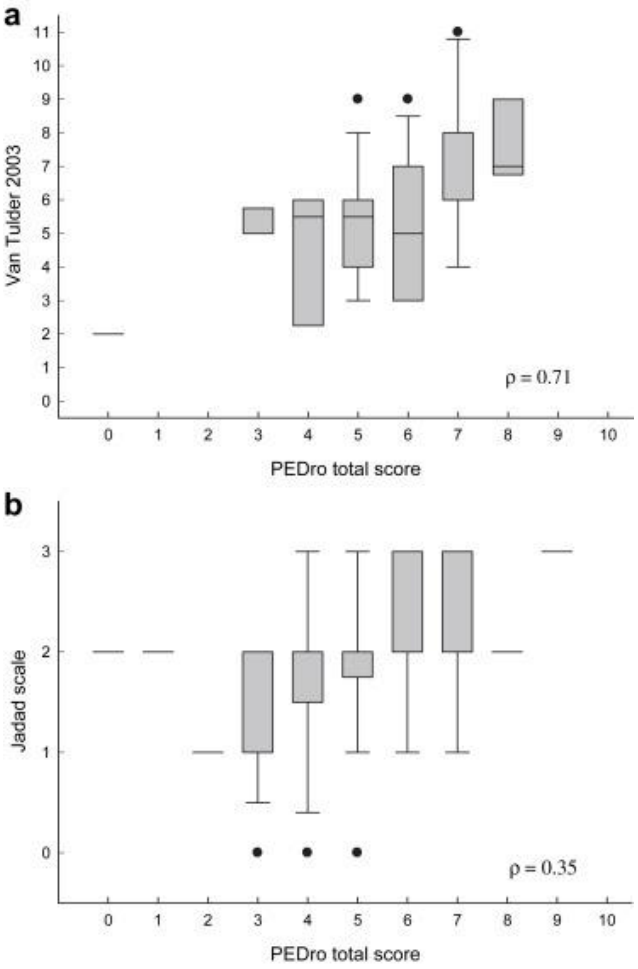
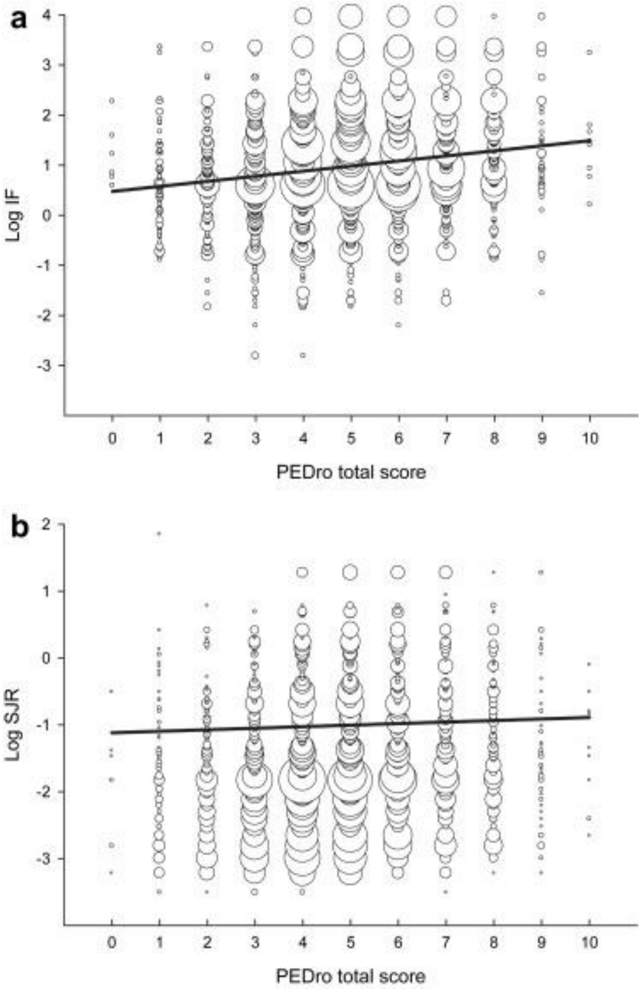


Image 2:



DISCUSSION:

The convergent besides construct rationality of Pedro total score are both supported by the findings of this research. In addition, we discovered evidence of concept validity for nine of the twelve questions that were used to calculate the overall score on Pedro [15]. It should not come as a surprise that there is only a modest link between the overall score on the Pedro test and the Jawad scale when convergent validity is considered [16]. It was reasonable to anticipate that there would be a modest association between the overall scores on these two measures given that two scales had just three substances in common [17]. On other side, researchers anticipated and found that there was a larger association between the Pedro and the Van Tudor scales due to the fact that the majority of the elements on these scales are shared by both scales [18].

When analyzing these data, there is one more factor that must be taken into account, and that is the disparity in the scoring methods used by various scales. When three distinct measures were used to evaluate the superiority of randomized measured tests, it was discovered that the scales' criteria and definitions were significantly varied from one another [19]. This resulted in different items being given different scores on three separate scales. This difficulty is also apparent in the data that we have collected for ourselves since there is only a minimal association between scores on the Jawad scale and scores on subscale of Pedro substances that are included in the Jawad scale [20]. Our current research demonstrates probable advantage that may be gained by increasing the uniformity of scale scoring systems [21].

Based on the data we have; it seems that the Pedro total score is able to differentiate among physiotherapy trials of better and poorer quality [22-29]. It would seem that there is not much of a gap between the 11th and 90th percentiles in terms of the IF of trials with Pedro's total scores, but there really is. However, the IF that is found in the middle of all of the experiments that are part of the Pedro database is 3.470, and the interquartile range is 1.700e5.150. In light of the aforementioned, we consider an influence of 2,200 IF points to be fairly significant [30]. Prior to the execution of this research project, there are only two studies published on the content validity of scales that were utilized to evaluate the quality of randomized clinical tests. These reports concerned the Yates scale and Jawad scale. In each of this exploration, the internal consistency was investigated by determining not if the total scores have been able discern among randomized clinical trials which were considered to be of exceptional quality among specialists and healthcare professionals (36 tests for the Jawad scale and 25 tests for the Yates scale) [31].

Researchers decided to employ a different technique since bibliometric indices of effect made it possible for us to include a far higher number of trials in the research, and they also give what is arguably a more objective criterion. For these reasons, we opted to utilize them [32]. Obviously, the procedural excellence remained evaluated by means of an indirect way, but unfortunately, our current item of constraint remains rather typical in corroboration studies since there is often no gold standard metric to use. Due to the fact that there are several possible alternate trials of construct validity, researchers do not consider the results of our research to be conclusive. Because of this, it is important to note that the results of this research should be considered preliminary [33].

There are other known or suspected factors that affect publication of a document in the journal (for example, positive outcomes remain extra probable to remain published through some journals); however, in over-all, higher-effect journals must still publish recovering tests. The construct that was selected to trial methodological quality is incomplete since there are additional known or suspected aspects that impact our current publication of the manuscript in the journal [34]. This presumption is given further credence by the fact that more influential publications insist on trial registration and the use of the CONSORT declaration throughout the peer-review process [35]. This is obvious that there are nonmethodological elements connected through publishing of the trial in a journal as well, in addition it can clarify why our research only showed modest connections between the two variables. The limited dispersion of Pedro ratings as well as the bibliometric measures of effect may have had a role in the weak connections that were discovered [36].

The results of the examination of each item that makes up the Pedro scale provide evidence in favor of the theory that existence of certain methodological traits is related to publication in journals that have a greater impact [37]. The intention-to-treat analysis had the highest connection with IF; research that satisfied this item had impact factors that were, on average, 1.16 impact factor points higher than research that did not fulfill our current item. The writing of point estimates and variability had the highest correlation with SJR scores; studies that complied with this item had JSRs that were, on median, 0.05 points higher than researches that did not comply. Because outcomes of logistic regression remained comparable to those of the linear regression, this finding lends support to hypothesis that the existence of personal characteristics is connected to better quality trials [38]. Analyses using univariate linear and logistic regression remained carried out; as a consequence, the study did not take into account any potential confounders influences, including those that may have been the consequence of other substances on Pedro scale.

When thinking about the findings of this research, there are a few caveats that need to be taken into consideration. The first difference is that we employed bibliometric directories for 2018, while tests remained published over the course of several years, during which time bibliometric indexes evolved [39]. This would lead to a reduction in the importance given to the evidence supporting construct validity. The fact that journal influence remains connected to, but not similar as experimental quality is the subject of the second constraint. Because of this, we did not anticipate finding substantial connections between the effect and the quality of the findings. Construct validity, as opposed to convergent validity or contrast through a predetermined gold standard, remains the kind of validity that is relied on in the majority of research, making this an issue that is widespread [40].

CONCLUSION:

There really is empirical evidence to support both convergent and composite reliability of Pedro over-all score, as well as internal consistency of nine out of the ten components that underwrite to Pedro over-all score.

REFERENCES:

1. Hoy D, March L, Woolf A, Blyth F, Brooks P, Smith E, et al. The global burden of neck pain: estimates from the global burden of disease 2010 study. Ann Rheum Dis. 2019;73(7):1309–15. Epub 2014/02/01. pmid:24482302
2. Cross M, Smith E, Hoy D, Carmona L, Wolfe F, Vos T, et al. The global burden of rheumatoid arthritis: estimates from the global burden of disease 2020 study. Ann Rheum Dis. 2014;73(7):1316–22. Epub 2020/02/20. pmid:24550173
3. Global Burden of Disease Study 2013 Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet. 2019;386(9995):743–800. Epub 2015/06/13. pmid:26063472
4. Hurwitz EL, Randhawa K, Yu H, Côté P, Haldeman S. The Global Spine Care Initiative: a summary of the global burden of low back and neck pain studies. Eur Spine J. 2018;27(Suppl 6):796–801. Epub 2020/02/27. pmid:29480409**.**
5. Baxter GD, Chapple C, Ellis R, Hill J, Liu L, Mani R, et al. Six things you need to know about low back pain. J Prim Health Care. 2020;12(3):195–8. pmid:32988440
6. Knoop J, van Lankveld W, Geerdink FJB, Soer R, Staal JB. Use and perceived added value of patient-reported measurement instruments by physiotherapists treating acute low back pain: a survey study among Dutch physiotherapists. BMC Musculoskelet Disord. 2020;21(1):120. pmid:32093706
7. Hemingway H, Croft P, Perel P, Hayden JA, Abrams K, Timmis A, et al. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. BMJ. 2019;346:e5595. Epub 2013/02/07. pmid:23386360
8. Hill JC, Dunn KM, Lewis M, Mullis R, Main CJ, Foster NE, et al. A primary care back pain screening tool: identifying patient subgroups for initial treatment. Arthritis Rheum. 2018;59(5):632–41. Epub 2008/04/29. pmid:18438893
9. Hill JC, Whitehurst DG, Lewis M, Bryan S, Dunn KM, Foster NE, et al. Comparison of stratified primary care management for low back pain with current best practice (STarT Back): a randomised controlled trial. Lancet. 2018;378(9802):1560–71. Epub 2011/10/04. pmid:21963002
10. Karstens S, Krug K, Raspe H, Wunderlich M, Hochheim M, Joos S, et al. Prognostic ability of the German version of the STarT Back tool: analysis of 12-month follow-up data from a randomized controlled trial. BMC Musculoskelet Disord. 2019;20(1):94. pmid:30819162

11. Foster NE, Mullis R, Hill JC, Lewis M, Whitehurst DG, Doyle C, et al. Effect of stratified care for low back pain in family practice (IMPaCT Back): a prospective population-based sequential comparison. Ann Fam Med. 2014;12(2):102–11. pmid:24615305

12. Sowden G, Hill JC, Morso L, Louw Q, Foster NE. Advancing practice for back pain through stratified care (STarT Back). Braz J Phys Ther. 2018;22(4):255–64. Epub 2018/07/05. pmid:29970301

13. Hsu C, Evers S, Balderson BH, Sherman KJ, Foster NE, Estlin K, et al. Adaptation and Implementation of the STarT Back Risk Stratification Strategy in a US Health Care Organization: A Process Evaluation. Pain Med. 2019;20(6):1105–19. Epub 2018/10/03. pmid:30272177

14. Campbell P, Hill JC, Protheroe J, Afolabi EK, Lewis M, Beardmore R, et al. Keele Aches and Pains Study protocol: validity, acceptability, and feasibility of the Keele STarT MSK tool for subgrouping musculoskeletal patients in primary care. J Pain Res. 2019;9:807–18. Epub 2016/10/30. pmid:27789972

15. Dunn KM, Campbell P, Lewis M, Hill JC, van der Windt DA, Afolabi E, et al. Refinement and validation of a tool for stratifying patients with musculoskeletal pain. Eur J Pain. 2021;25(10):2081–93. Epub 2021/06/09. pmid:34101299

16. Burgess R, Mansell G, Bishop A, Lewis M, Hill J. Predictors of Functional Outcome in Musculoskeletal Healthcare: An Umbrella Review. Eur J Pain. 2019;24(1):51–70. Epub 2019/09/12. pmid:31509625

17. van den Broek AG, Kloek CJJ, Pisters MF, Veenhof C. Validity and reliability of the Dutch STarT MSK tool in patients with musculoskeletal pain in primary care physiotherapy. PLoS One. 2021;16(3):e0248616. pmid:33735303

18. Dunn KM, Campbell P, Afolabi EK, Lewis M, van der Windt D, Hill JC, et al. Refinement and Validation of the Keele STarT MSK Tool for Musculoskeletal Pain in Primary Care. Rheumatology. 2017;56(suppl_2).

19. Beaudart C, Criscenzo L, Demoulin C, Bornheim S, van Beveren J, Kaux J-F. French translation and validation of the Keele STarT MSK Tool. European Rehabilitation Journal. 2021;1(1):1–7.

20. Rysstad T, Grotle M, Aasdahl L, Hill JC, Dunn KM, Tingulstad A, et al. Stratifying workers on sick leave due to musculoskeletal pain: translation, cross-cultural adaptation and construct validity of the Norwegian Keele STarT MSK tool. Scandinavian Journal of Pain. 2022. pmid:35148473

21. Ben Ami N, Hill J, Pincus T. STarT MSK tool: Translation, adaptation and validation in Hebrew. Musculoskeletal care. 2021;n/a(n/a). pmid:34862708

22. Hay EM, Dunn KM, Hill JC, Lewis M, Mason EE, Konstantinou K, et al. A randomised clinical trial of subgrouping and targeted treatment for low back pain compared with best current care. The STarT Back Trial Study Protocol. BMC Musculoskelet Disord. 2008;9:58. Epub 2008/04/24. pmid:18430242

23. Hingorani AD, Windt DA, Riley RD, Abrams K, Moons KG, Steyerberg EW, et al. Prognosis research strategy (PROGRESS) 4: stratified medicine research. BMJ. 2013;346:e5793. Epub 2013/02/07. pmid:23386361

24. Protheroe J, Saunders B, Bartlam B, Dunn KM, Cooper V, Campbell P, et al. Matching treatment options for risk sub-groups in musculoskeletal pain: a consensus groups study. BMC Musculoskelet Disord. 2019;20(1):271. Epub 2019/06/04. pmid:31153364

25. Sowden G, Hill JC, Konstantinou K, Khanna M, Main CJ, Salmon P, et al. Targeted treatment in primary care for low back pain: the treatment system and clinical training programmes used in the IMPaCT Back study (ISRCTN 55174281). Fam Pract. 2012;29(1):50–62. pmid:21708984

26. Saunders B, Hill JC, Foster NE, Cooper V, Protheroe J, Chudyk A, et al. Stratified primary care versus non-stratified care for musculoskeletal pain: qualitative findings from the STarT MSK feasibility and pilot cluster randomized controlled trial. BMC Fam Pract. 2020;21(1):31. pmid:32046656
27. Corp N, Mansell G, Stynes S, Wynne-Jones G, Morsø L, Hill JC, et al. Evidence-based treatment recommendations for neck and low back pain across Europe: A systematic review of guidelines. European Journal of Pain. 2021;25(2):275–95. pmid:33064878
28. Back-UP. The Back-UP web app demonstration for clinicians 2020. Available from: https://www.youtube.com/watch?v=kgyodFFHEJ8.
29. SoSci Survey GmbH. SoSci Survey–the Solution for Professional Online Questionnaires o.J. [024.03.2021]. Available from: https://www.soscisurvey.de/en/index.
30. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. COSMIN checklist manual COSMIN initiative; 2012. Available from: http://www.cosmin.nl/images/upload/files/COSMIN%20checklist%20manual%20v9.pdf.
31. Beaton D, Bombardier C, Guillemin F, Ferraz MB. Recommendations for the Cross-Cultural Adaptation of the DASH & QuickDASH Outcome Measures: Institute for Work & Health; 2007 [04.03.2022]. Available from: http://dash.iwh.on.ca/sites/dash/files/downloads/cross_cultural_adaptation_2007.pdf.
32. Aebischer B, Hill JC, Hilfiker R, Karstens S. German translation and cross-cultural adaptation of the STarT Back Screening Tool. PLoS One. 2015;10(7):e0132068. pmid:26161669
33. Mahler C, Rochon J, Karstens S, Szecsenyi J, Hermann K. Internal consistency of the readiness for interprofessional learning scale in German health care students and professionals. BMC Med Educ. 2014;14(1):145. Epub 2014/07/17. pmid:25027384
34. Guss CD. What Is Going Through Your Mind? Thinking Aloud as a Method in Cross-Cultural Psychology. Front Psychol. 2018;9:1292. Epub 2018/08/29. pmid:30150948
35. Schmidt CO, Kohlmann T, Pfingsten M, Lindena G, Marnitz U, Pfeifer K, et al. Construct and predictive validity of the German Orebro questionnaire short form for psychosocial risk factor screening of patients with low back pain. Eur Spine J. 2016;25(1):325–32. Epub 2015/08/28. pmid:26310842
36. Stadler M, Sailer M, Fischer F. Knowledge as a formative construct: A good alpha is not always better. New Ideas Psychol. 2021;60:100832.
37. Cramer H, Lauche R, Langhorst J, Dobos GJ, Michalsen A. Validation of the German version of the Neck Disability Index (NDI). BMC Musculoskelet Disord. 2014;15:91. Epub 2014/03/20. pmid:24642209
38. Angst F, Goldhahn J, Pap G, Mannion AF, Roach KE, Siebertz D, et al. Cross-cultural adaptation, reliability and validity of the German Shoulder Pain and Disability Index (SPADI). Rheumatology (Oxford). 2007;46(1):87–92. Epub 2006/05/25. pmid:16720638
39. Exner V, Keel P. [Measuring disability of patients with low-back pain—validation of a German version of the Roland & Morris disability questionnaire]. Schmerz. 2000;14(6):392–400. Epub 2003/06/12. pmid:12800012
40. Stucki G, Meier D, Stucki S, Michel BA, Tyndall AG, Dick W, et al. [Evaluation of a German version of WOMAC (Western Ontario and McMaster Universities) Arthrosis Index]. Z Rheumatol. 1996;55(1):40–9. Epub 1996/01/01. pmid:8868149.