



# Patient Identified Problem (PIP) Scale, Validity, Reliability, Responsiveness, Likelihood ratio, and Minimal Clinically Important Difference

Adi Halili

\*Correspondence: [halilipt@msn.com](mailto:halilipt@msn.com)



Halili Physical Therapy, 268 E River rd #130, Tucson AZ 85704, USA.

## Abstract

**Background:** Patient identified problems (PIP) are a component of hypothesis oriented algorithm for clinician (HOAC) model.

**Objective:** This study evaluated the statistical properties of an outcome tool (PIP scale) developed from the PIP in the model.

**Design:** Observational retrospective.

**Methods:** Blinded records were used to measure change in the PIP scale and individual problem scores in a patient population that did not receive treatment and patients that received treatment. The analysis included measurements of construct and concurrent validity, reliability, responsiveness including AUC, likelihood ratios, specificity and sensitivity, and establishment of minimal clinical important difference (MCID).

**Results:** Construct validity was demonstrated by showing no significant change when no treatment was provided (avg change -0.59 (95% CI -1.8 to 0.6,  $p = 0.34$ ) and significant change when treatment was provided (avg change 14.46, (12.57 to 16.35  $p < 0.0001$ ) a weak/moderate positive correlation with ODI, NDI, DASH, and LEFS ( $r = 0.27, 0.41, 0.45, 0.30$  respectively) established a level of concurrent validity. Scale reliability was excellent (ICC = 0.96, 95% CI 0.93 to 0.97). Excellent Responsiveness was demonstrated by AUC 0.78, +LR 7.55, -LR 0.39, specificity 91.46, and sensitivity 64.45. MCID was determined to be 3.8 points (95% CI 1.4 to 8.2).

**Limitations:** The validity of this study was established under specific conditions where reporting was done several days removed from treatment, using self-recording dialogue, and with the prior scores provided. Findings need to be replicated in future studies.

**Conclusions:** This study demonstrated that the PIP scale is a simple, versatile tool and has good validity both for day to day clinical use and large-scale research.

**Keywords:** Outcome measures, HOAC model, rehabilitation, physical therapy, back pain, neurological rehabilitation

## Background

The purpose of this study was to evaluate the statistical validity of an outcome tool defined as the "Patient Identified Problem" scale or PIP scale in short. Being a physical therapist in a neurologically based outpatient setting, one encounters many patients with a complex and confusing clinical presentation. To treat

this population successfully, three areas of focus are identified:

1. Use an evaluation method that accounts for the clinical complexity.
2. Standardization of the treatment approach
3. A valid outcome measure tool or tools that can evaluate overall progress as well as changes in the individual

co-ponents of the clinical presentation.

The hypothesis oriented algorithm for clinician (HOAC) model was initially developed by Rothstein in 1986 and further expanded upon in 2003 [1,2].

It provides a proper evaluation and assessment option. The use of this model is desirable because it allows for the initiation of treatment despite a certain level of initial uncertainty. In practice, this model is used in the following manner: after an extensive patient interview, a list of patient’s identified problems (PIP) is created and graded using the scale discussed in this paper. Several diagnostic hypotheses are developed. If needed, additional diagnostic tests and measurements are done, and a plan of care is established. Each visit, based on the patient response, the physical therapist evaluates if the ongoing diagnostic hypotheses are still plausible and would add additional hypotheses if indicated. The plan of care would also be modified accordingly.

The second issue outlined previously regarding standardization of care is addressed by the formation of several dozen treatment protocols and protocol sequences. A detailed description of this protocol system is discussed elsewhere [14] and is mostly beyond the scope of this paper.

Although there are several existing validated outcome tools already available, none provides an adequate answer to the challenge of evaluating the progress of a patient with a complicated clinical presentation. For example, Oswestry Disability Index (ODI) [5], Neck Disability Index (NDI) [6], Disability Arm Hand and Shoulder Index (DASH) [7], or Lower Extremity Functional Scale (LEFS) [8] are intended to measure a specific body region or part and do not provide information about other specific problems as well as the overall clinical status. While the Global Rating of Change scale (GRoC) [17] can provide information about overall progress, it does not provide problem-specific information. Finally, tools such as the Patient Specific Functional Scale (PSFS) [3] and pain rating scales [4] do provide specific information about some problems but are limited to specific criteria such as function or pain.

On the other hand, the PIP scale, if validated, can provide information about any specific problem identified by the physical therapist or the patient. It can provide information about a possible relationships between problems and the treatment provided as well as an overall measurement of change.

The PIP scale is a 1 to 10 (half point permitted) scale. The patient can score between a “1” to denotes that the problem is not currently active and “10,” which denotes maximal intensity. The problems are looked at both individually and as a cumulative number. The cumulative number is calculated according to the following formula:

$$SUM (individual\ problems\ score/number\ of\ problems) \times 10$$

(add the scores of all the individual problems, divide the number by the number of individual problems then multiply this number by 10).

Development of the individual set of problems and initial scoring is done during the initial evaluation. Although the term used is “Patient Identified Problem” or PIP, the problems selected for monitoring include both PIP and “Non-Patient Identified Problems” (NPIP) identified initially in the HOAC model. The reason for converging the two is that all problems are chosen in consultation with the patient, and once a problem is identified, it does not appear to matter if the patient or the physical therapist originates it. To ensure that the list of problems is similar to the actual range of the patient’s impairments, the initial evaluation must include a thorough patient interview and physical exam that include a complete systems evaluation. This manner of evaluation is the standard of care in allopathic medicine, and as physical therapy matures to a doctoral level profession, it must assume this standard as well. For example, if a patient is presenting with the chief complaint of right knee pain but also identifies that they have difficulty sleeping, constipation, anxiety, and sensitivity to cold, all items need to be included on the problem list.

Other procedural considerations done to decrease bias and improve ease of use include:

1. After the initial evaluation, the patient always records the numbers on the next visit and not immediately after treatment.
2. When possible, recording the numbers is done in a self-administrated manner on a computer dedicated for check-in before the visit
3. During check-in, the patient is provided with a list of the prior numbers (Figure 1). Because this scale is focusing on the measurement of change, providing the prior numbers is intended to reduce measurement error when the

**Patient Identified Problem Form (PIP)**

Below please provide detailed information of the problem. If this is a returning visit, please indicate what was better since the last visit, what is still a problem. Also note any other occurrences that might have influenced your symptoms (such as treatment by other practitioner, medical procedures or re-injury events)

Please list/update areas of problems, enter or change the problem score based on how you did this week, use a scale of 10 (worse) to 1 (no problem) for all items. Remember, what you are grading is how much this issue is interfering with your life not necessarily just current levels of pain

Ability to drive	2
Ability to get on a plane	4
Anxiety	3
Congestion right leg	2
IBS	2
Near fainting episodes	1
Neck pain	2
Pain in left eye	2
Vertigo	2

Figure 1. Login page.

- baseline number is not recalled correctly by the patient. The notion that providing the prior numbers decrease bias is supported in the literature by Guyatt et al [17].
4. When reporting a problem involving pain, the patient is to report it in terms of problem severity, not in terms of pain intensity. This allows using a single number regardless if the pain is intermittent or fluctuates in intensity.
  5. In addition to recording the scores, during the check-in process or while talking to the physical therapist, the patient can enter a subjective statement as well (Figure 1). Although not officially part of the scale, the subjective statement adds a qualitative dimension to the quantitative information provided by the PIP scale. This concept allows the clinician to adhere to the HOAC algorithm by providing additional information when trying to decide whether to continue with the original plan or make changes.
  6. When needed, additional problems can be added by the physical therapist and scored by the patient. If a problem is no longer active, other than being taken off the list, it is being scored as a "1," so it is available for the patient to see if the problem became active in the future.
  7. In addition to doing the PIP scale each visit, every several visits, most patients take a standardized tool such as the ODI [5], NDI [6], DASH [7], or LEFS [8]. The use of these scales was historically done in lieu of prior validation of the PIP scale, but these scales can still provide some additional information about the patient status.
  8. All information, including scores, subjective statements, and treatment provided, was automatically recorded in a Microsoft Access database allowing for further systematic analysis.

The following example illustrates how the scale is used in clinical practice:

During the initial evaluation, the patient in this case presented with nine problems: ability to drive, ability to get on a plane, pain in the left eye, near fainting episodes, neck pain, congestion right leg, anxiety, IBS, and vertigo. Upon completion of the evaluation, the therapist developed several complementary or competing hypotheses for the differential diagnosis. These were: sensitized status, neuritis, and circulatory congestion. The treatment plan included the practice standard desensitization protocol sequence: UD, DCS, Barral, CCCV, SYMPN followed by the lower extremity decongestive protocol sequence and additional protocols to address the neuritis causing the eye pain and vertigo. A more specific discussion about the protocols is available elsewhere [14]. A few examples to demonstrate how the PIP scale information is used are included next. Figure 2 is a condensed progress report that includes a subjective statement, The treatment provided, the number of total and active PIP problems (scored above 1), The PIP scale cumulative score, other scale used (none in this case), and other scale scores if available. Please note that this is only a portion of the actual treatment record, and it is not intended to represent the full physical therapy documentation

for this encounter. The following example includes the first five visits, which included the five protocols currently used to address the central sensitization (see previous comments). With a glance, the therapist can see that the PIP scale dropped by 17 points from 59 to 42. Together with the information on the subjective statements, it is reasonable to infer that the patient is responding to the treatment predictably and that so far, the original diagnostic hypotheses still stand.

However, if there are still questions as to what problems are getting better and which ones still need to be addressed, the therapist can refer to another summary report (Figure 3). This summary (which in this case includes the whole episode of care) includes numeric information on each problem. This information can be used to further understand the response of a specific problem to the intervention as well as possible relationships between the problems. For example, while anxiety, IBS, vertigo, and near-fainting episodes, are expected to drop after the desensitization sequence, the drop in neck pain and congestion in the leg were less obvious. Alternatively, improvement in ability to drive did not occur until the neuritis in the eye had started to improve. These insights are an essential part of helping in the formation of better understanding of both the pathology and mechanisms of the intervention.

Although this information is essential in a practice setting that treats patients with complicated clinical presentation, the ability of this tool to provide multi-dimensional layers of information can be utilized in any practice setting.

The purpose of this study is to evaluate the statistical properties of the PIP scale.

## Methods

Using a new Microsoft Access database file, data was imported from the primary database, including records from April 1, 2015, until December 31, 2018, stripping all identifiable information. Study database included 1466 patients (956 Female, 510 Male, age range from eight months to 97 calculated from the study mid-point in 2017, avg age 61, std 15.9) with a total of 18,747 treatment visits.

From this population, seven samples were taken: "Treatment Group", "No Treatment Group", "7 Day No Treatment", "PIP/single-problem/ODI", "PIP/NDI", "PIP/DASH", and "PIP/LEFS". Each sample included a pair of two consecutive measurements yielding the change in score that was analyzed in this study. All samples were similar in female/male distribution, average age, and age distribution. Welch-test for unequal variances was used to evaluate *p* scores and confidence interval (CI). Table 1 summarizes the sample characteristics and the manner they were used in this study.

To create a "Treatment Group," 660 patients with 12,428 visits were isolated (445 female, 215 male, age range 11 to 97 avg age 58 std 17.89). Filtering was done by searching individual PIPs, using the keywords "Lumbar," "Back," and "Pelvic," meaning that at least one of their PIP included back, lumbar, or pelvic pain.

progress report

#PIP	active	PIPScale	OS	OS score
9	9	59		

6/5/2019  
 65 yo female with cc patient is blind in left eye in 2016 had T bone MVA since has anxiety about driving after some chiropractic treatment able to drive locally only about 6 weeks ago developed vertigo after doing some exercises but this is mostly resolved also have sensation of tightness left leg vision in left eye was lost 7 years ago possibly due to infection of ischemic event. Still a source of pain some neck tightness had couple of fainting episodes also episode in 2014 recently had a headache but no regular headaches  
 Rx UD  
 Treatment response (recorded the following encounter):

#PIP	active	PIPScale	OS	OS score
9	8	49	NONE	

6/14/2019  
 patient reported somewhat less anxiety this week also less congestion in the leg  
 Rx DCS  
 Treatment response (recorded the following encounter):

#PIP	active	PIPScale	OS	OS score
9	9	56	NONE	

6/24/2019  
 Patient reports that this week she had at least 3 episodes of vertigo. No headaches since she started the treatments here.  
 Rx Barral  
 Treatment response (recorded the following encounter):

#PIP	active	PIPScale	OS	OS score
9	9	52	NONE	

7/1/2019  
 patient reports able to drive locally with less stress  
 Rx CCCV  
 Treatment response (recorded the following encounter):

#PIP	active	PIPScale	OS	OS score
9	7	44	NONE	

7/11/2019  
 patient reported had couple of days of adjustment after the last treatment blood pressure was high then better, had difficulty with deep breathing  
 Rx SYMPN  
 Treatment response (recorded the following encounter):

#PIP	active	PIPScale	OS	OS score
9	7	42	NONE	

Figure 2. Progress summary form.

Problem Description	pip progression table																	
	6/5	6/14	6/24	7/1	7/11	7/16	7/26	8/2	8/9	8/15	8/27	9/3	9/10	9/24	10/4	10/11	10/17	1/7
ability to drive	8	8	6	6	7	7	6	6	6	6	7	7	7	7	7	2	2	2
Ability to get on a plane	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	6
anxiety	6	6	6	5	5	4	4	2	2	2	2	2	3	3	3	3	3	3
congestion right leg	6	4	4	3	4	2	3	2	1	2	3	2	3	3	3	3	3	2
IBS	5	5	5	5	2	7	5	5	4	1	1	7	7	5	4	4	4	2
near fainting episodes	5	5	5	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
neck pain	5	3	4	4	1	1	3	3	3	2	2	1	3	4	3	2	2	2
pain in left eye	3	2	2	5	7	4	5	7	7	7	4	7	4	2	2	2	5	2
vertigo	5	1	8	6	3	2	2	2	2	2	1	1	1	1	5	3	4	2

Figure 3. Individual PIP summary table.

**Table 1. Sample characteristics.**

	Sample name	Query terms	Size	The manner used for in this study	Demographics	Test for <i>p</i> and CI
1	Study population	“Between April 1, 2015, and December 31, 2018,”	1466	Source	956 F, 510 M, Age Avg:61 Range 8 months to 97, std 15.9	NA
2	Treatment group	“Lumbar,” “Back” and “Pelvic”	660	Construct validity, Responsiveness, MCID	445 F, 215 M, avg age 58, range 11 to 97 std 17.89	Welch-test unequal variances
3	No treatment group	“No Treatment”	65 (82 occurrences, 610 scores)	Construct validity, Reliability Responsiveness MCID	48 F, 17 M, avg age 63 range 13 to 97 std 16.1)	Welch-test unequal variances
4	Seven days no treatment	“No Treatment” and Visit Date – Next Visit <7	49	Construct validity, Reliability Responsiveness	34 F, 15 M, Avg age 61, range 13 to 87 std 17	Welch-test unequal variances
5	PIP/single-problem/ODI	“ODI”	289 (686 PIP, 681 ind. problem)	Concurrent validity	201 F, 88 M, avg age 60, range 13-94 std 16	Welch-test unequal variances
6	PIP/NDI	“NDI”	520	Concurrent validity	436 F 84 M avg age 59, range 18-83 std 12.6	Welch-test unequal variances
7	PIP/DASH	“DASH”	223	Concurrent validity	151 F 72 M avg age 59 range 11-86 std 16.6	Welch-test unequal variances
8	PIP/LEFS	“LEFS”	407	Concurrent validity	288 F, 119 M avg age 62 range 12-91 std 17.9	Welch-test unequal variances

The “No Treatment Group” includes record pairs with a patient sign-in on a day that no treatment was provided, coupled with an available subsequent visit score. This pairing of two records without treatment provided because, at times, when patients could not stay for their appointment after login due to the therapist being late or other reasons. The “No Treatment Group” included 65 patients (82 occurrences of sign-in without treatment coupled with a subsequent sign-in sometime later) (48 female, 17 male, age range 13 to 97 avg 63 std 16.1). The average number of days between sign-ins was 17.6 (95% CI 15.2 to 20.0). This category included 611 individual problems. A subgroup was also created to include 49 patients (49 occurrences) (34 female, 15 male, age range 13 to 87 Avg 61 std 17) with 415 individual problems where the follow-up visit occurred seven days or less after the no treatment visit. This group is called the “7 Day No Treatment” (average number of days between sign-in 4.5, 95% CI 4.2 to 4.7). The validation process of the individual scores was done using both control groups. The reason both groups were used is the theory that the “7 Day No Treatment” group could provide a better measure of the measurement method due to the shorter time between visits, while the generic no treatment group offered a more accurate measure of a treatment effect over other non-treatment related changes that can occur during that period. However, in retrospect, no significant difference was detected between the two non-treatment groups (avg change -0.023,  $p = 0.72$ , 95% CI -0.15 to 0.10).

To demonstrate construct validity, a paired two samples t-test (Welch) of either the “No Treatment” or “7 Day No

Treatment” was done to demonstrate that the mean score is consistent with zero.

Also, an independent samples t-test was done to evaluate if there is a significant difference between the “No Treatment” or the “7 Day No Treatment” group and a “Treatment Group” PIP and single score changes across an episode of care (defined as the difference between the first available score in the study period and the last available score in the study period).

**Concurrent validity** was evaluated using paired samples of PIP scale and ODI, NDI, DASH, and LEFS, testing the correlation of the change in the scales scores between two visits.

Also, a change in Individual score of back pain problem was correlated with ODI score change, when the pairing was available, to assess for concurrent validity between a single complaint of back pain and ODI.

**Reliability** was tested using an Intraclass correlation coefficient (ICC) with 95% CI of two consecutive tests in the “No Treatment” group, “7 Day No Treatment” group for an individual problem, and the PIP “No Treatment” group for the PIP scale.

**Responsiveness** was evaluated by plotting receiver operator characteristic (ROC) curves. The area under the curve (AUC) was plotted to indicate the accuracy of classifying patients in the “No Treatment”, “7 Day No Treatment” group, or the “Treatment Group”, The Youden index *J*, specificity scores, sensitivity scores positive and negative likelihood ratios (LR) were also calculated to evaluate the scale responsiveness level.

**Minimal clinically important difference (MCID)** and its confidence-interval were evaluated using the one-half standard deviation method [10]. The measurement was taken from

the std of the mean change between the treatment and no treatment groups. This measurement was taken for both for an individual problem and the PIP scale.

Statistical analysis was done using MedCalc Software [9].

## Results

**Table 2** summarizes the PIP scale and individual score construct and concurrent validity, reliability, and responsiveness, and MCID.

In this study, **construct validity** measured the degree of accuracy measuring no change when there is none and measuring change when there is. In this case, construct validity was demonstrated because the three samples that were supposed to measure no change did just that, and the two samples supposing to measure change did so as well.

Correlation between the PIP and other scales was positive but with only weak to moderate strength. While both PIP and the other scales are more likely than not to identify positive and negative changes together, the less than strong correlation could be because that the other scales only measure a fraction of the problems measured in the PIP scale, but further studies are needed to answer this question.

**Reliability** was measured by the ability of a repeated measurement to produce the same results. All three sample groups used in this study demonstrated excellent reliability.

**Responsiveness** was demonstrated using several derivatives of **specificity** and **sensitivity** measurements, including AUC, LR, and Youden index J. It was felt that the use of several of these could facilitate easier comparisons in future studies.

**Table 3** provides some PIP scale comparison data with other standardized scales.

The **MCID** is probably the most clinically useful data for this scale since it can allow the clinician to evaluate expeditiously if a numeric change on the scale is meaningful or not. **Table 3** also provides a comparison of the MCID data.

## Discussion

As discussed in the introduction, the PIP scale provides the clinician with information not available in other similar self-reporting scales such as ODI, NDI, DASH, LEFS, PSFS, VAS, and GRoC. The example discussed in the introduction provides a succinct example of this assertion. The PIP scale in that example included nine problems: ability to drive, ability to get on a plane, pain in the left eye, near fainting episodes, neck pain, congestion right leg, anxiety, IBS, and vertigo. While the other scales can evaluate some of the problems, for example, NDI (neck pain), PSFS (ability to get on a plane, ability to drive), LEFS (congestion right leg), VAS (neck pain, pain in left eye), none of these evaluated the full list of problems. This observation can help explain why the correlation between these other scales and the PIP scale was less than expected: they only partially measure the set of problems; hence the results are only partially correlated.

An additional feature of the PIP scale not available on the

other scales is the ability to easily observe the relationships between individual problems as the episode of care is progressing.

The final item that makes this scale useful, not only in a practice that treats patients with a complicated neurological clinical presentation but also in any orthopedic based outpatient practice setting, is the ease of use. The PIP scale is a self-administrated, easily computerized tool. Patients can enter the data in a few minutes and provide the clinician with immediate valuable information before the therapy session begins.

Without understanding the statistical behavior of this scale, however, the value of this tool is limited. Therefore, the results of the analysis done in this study are a valuable first step in this direction.

At the root of all self-reporting tools, what is measured is the ability of the intact human brain to detect change. As such, it was expected that this scale would exhibit a statistical behavior with similar characteristics to the other scales used as anchors in this study, and for the most part, this was the case.

**Construct validity** was established by demonstrating no statistically significant difference with repeated measurement of the control groups and a significant change with repeated measurement when treatment was provided. **The high-reliability** level was established by finding the ICC above 90%. **Responsiveness** is comparable to the other scales discussed. Responsiveness was established by measuring AUC, specificity and sensitivity, Youden index, a high positive likelihood ratio, a low negative likelihood ratio. A workable **MCID** score was established to assist with day to day interpretation.

However, there are a couple of issues that should be pointed out. First, when checking for **concurrent validity**, it was partially established by finding a positive correlation with all anchor scales used but with only a weak to moderate correlation. As discussed earlier, it is hypothesized that this discrepancy can be explained by the larger number of issues measured in the PIP compared to the other scales, but more study is needed. The other difference was found in the **specificity** and **sensitivity** levels. While specificity levels were higher than all the other scales compared, sensitivity was lower. The possible meaning of this is that the PIP scale is less likely to show change, either when there is none (high specificity) and where there is one (low sensitivity). Also, when using the MCID to evaluate if a change had occurred, one should consider the confidence interval associated with this number (3.8 95% CI 1.4 to 8.2). For example, if a change of seven points was recorded on the PIP scale, caution should be taken, interpreting this number as a clinically important difference because it is still smaller than the high end of the CI. Finally, one must remember that this study is the first time the tool was assessed, and additional studies are warranted to validate this tool by replicating these findings.

With these caveats noted, this study does provide the much-needed initial statistical validation to this versatile and simple to use tool that can be integrated with the HOAC "Citation:" on

**Table 2. Scale Validation Results.**

<b>Construct Validity</b>					
<i>1. two measurements with no treatment in between, expected no significant change</i>					
	Size	Avg days between tests (95% CI)	Avg change (95% CI)	p-value	Comments
Individual score 7 Day No Treatment	415	4.5 (4.2 to 4.7)	-0.19 (-0.09 to 0.05)	0.62	No significant change
Individual score No Treatment	610	17.6 (15.2 to 20.0)	-0.04 (-0.14 to 0.05)	0.43	No significant change
PIP scale No Treatment	82	17.8 (11.0 to 24.6)	-0.59 (-1.8 to 0.6)	0.34	No significant change
<i>2. Two measurements with treatment in between, expected significant change</i>					
	Size No Rx, Rx	Avg days between tests (95% CI) no Rx, Rx	Avg change (95% CI)	p-value	Comments
Individual score No Treatment to Treatment	610, 660	17.6 (15.2 to 20.0), 17.5 (13.6 to 21.5)	1.39 (1.16 to 1.63)	<0.001	Significant change
PIP scale No Treatment to Treatment	82, 660	17.8 (11.0 to 24.6), 17.5 (13.6 to 21.5)	14.46 (12.57 to 16.35)	<0.001	Significant change
<b>Concurrent validity</b>					
<i>Correlation between PIP scale and other scales</i>					
Scale comparison	Sample size	Correlation coefficient r (95% CI for r)	p value	Comments	
ODI and PIP Scale	686	0.27 (0.20 to 0.34)	<0.001	Weak positive	
ODI and individual PIP for back pain	681	0.34 (0.27 to 0.40)	<0.001	Weak positive	
NDI and PIP Scale	520	0.41 (0.34 to 0.48)	<0.001	Moderate positive	
DASH and PIP Scale	223	0.45 (0.34 to 0.55)	<0.001	Moderate positive	
LEFS and PIP Scale	407	0.30 (0.21 to 0.39)	<0.001	Weak positive	
Reliability	Size	Avg days between tests (95% CI)	ICC (95% CI)	Comments	
Individual score 7 Day No Treatment	415	4.5 (4.2 to 4.7)	0.96 (0.95 to 0.97)	Excellent reliability	
individual score No Treatment	610	17.6 (15.2 to 20.0)	0.91 (0.89 to 0.92)	Excellent reliability	
PIP scale No Treatment	82	17.8 (11.0 to 24.6)	0.96 (0.93 to 0.97)	Excellent reliability	
Responsiveness (from ROC curve)	AUC	+LR (95% CI), -LR (95%CI)	Specificity and sensitivity	Youden index J	MCID (1/2 SD)
PIP Scale	0.78	7.55 (3.7 to 15.4), 0.39 (0.3 to 0.4)	91.46 and 64.45	0.56	3.8 (CI 95% 1.4 to 8.2)
Individual score	0.68	4.38 (3.5 to 5.4), 0.5 (0.5 to 0.6)	87.21 to 55.98	0.43	0.89 (CI 95% 0.33 to 1.5)

**Table 3. Comparison between scales of Responsiveness and MCID.**

Scale	AUC	Specificity	Sensitivity	MCID
PIP	0.78	91.46%	64.45	3.8
ODI [12]	0.71	63%	76%	13.67
NDI [13]	0.96	81%	98%	7
DASH [11,15]	0.77 [15]	74% [11]	82% [11]	10.83 [11]
LEFS [8]	0.76	70%	81%	9

page 8 model or any other analytical assessment process.

**Competing interests**

The author declares that he has no competing interests.

**Publication history**

Editor: Mohammad H. Hadadzadeh, Wheeling Jesuit University, USA.  
 Received: 05-Apr-2020 Final Revised: 18-May-2020  
 Accepted: 28-May-2020 Published: 09-Jun-2020

## References

1. Rothstein JM and Echternach JL. **Hypothesis-oriented algorithm for clinicians. A method for evaluation and treatment planning.** *Phys Ther.* 1986; **66**:1388-94. | [Article](#) | [PubMed](#)
2. Rothstein JM, Echternach JL and Riddle DL. **The Hypothesis-Oriented Algorithm for Clinicians II (HOAC II): a guide for patient management.** *Phys Ther.* 2003; **83**:455-70. | [PubMed](#)
3. Hefford C, Abbott JH, Arnold R and Baxter GD. **The patient-specific functional scale: validity, reliability, and responsiveness in patients with upper extremity musculoskeletal problems.** *J Orthop Sports Phys Ther.* 2012; **42**:56-65. | [Article](#) | [PubMed](#)
4. Ware LJ, Epps CD, Herr K and Packard A. **Evaluation of the Revised Faces Pain Scale, Verbal Descriptor Scale, Numeric Rating Scale, and Iowa Pain Thermometer in older minority adults.** *Pain Manag Nurs.* 2006; **7**:117-25. | [Article](#) | [PubMed](#)
5. Fairbank JC and Pynsent PB. **The Oswestry Disability Index.** *Spine (Phila Pa 1976).* 2000; **25**:2940-52; discussion 2952. | [Article](#) | [PubMed](#)
6. Vernon H and Mior S. **The Neck Disability Index: a study of reliability and validity.** *J Manipulative Physiol Ther.* 1991; **14**:409-15. | [PubMed](#)
7. Polson K, Reid D, McNair PJ and Larmer P. **Responsiveness, minimal importance difference and minimal detectable change scores of the shortened disability arm shoulder hand (QuickDASH) questionnaire.** *Man Ther.* 2010; **15**:404-7. | [Article](#) | [PubMed](#)
8. Binkley JM, Stratford PW, Lott SA and Riddle DL. **The Lower Extremity Functional Scale (LEFS): scale development, measurement properties, and clinical application.** *North American Orthopaedic Rehabilitation Research Network.* *Phys Ther.* 1999; **79**:371-83. | [PubMed](#)
9. Schoonjans F. **MedCalc statistical software.** *MedCalc.* 2019. | [Website](#)
10. Farivar SS, Liu H and Hays RD. **Half standard deviation estimate of the minimally important difference in HRQOL scores?** *Expert Rev Pharmacoecon Outcomes Res.* 2004; **4**:515-23. | [Article](#) | [PubMed](#)
11. Franchignoni F, Vercelli S, Giordano A, Sartorio F, Bravini E and Ferriero G. **Minimal clinically important difference of the disabilities of the arm, shoulder and hand outcome measure (DASH) and its shortened version (QuickDASH).** *J Orthop Sports Phys Ther.* 2014; **44**:30-9. | [Article](#) | [PubMed](#)
12. Monticone M, Baiardi P, Vanti C, Ferrari S, Pillastrini P, Mugnai R and Foti C. **Responsiveness of the Oswestry Disability Index and the Roland Morris Disability Questionnaire in Italian subjects with sub-acute and chronic low back pain.** *Eur Spine J.* 2012; **21**:122-9. | [Article](#) | [PubMed](#) | [Abstract](#) | [PubMed FullText](#)
13. Monticone M, Ambrosini E, Vernon H, Brunati R, Rocca B, Foti C and Ferrante S. **Responsiveness and minimal important changes for the Neck Disability Index and the Neck Pain Disability Scale in Italian subjects with chronic neck pain.** *Eur Spine J.* 2015; **24**:2821-7. | [Article](#) | [PubMed](#)
14. Halili, A. **Systemic Manual Therapy.** Sun Bernardino, California: 2020, Kindle direct publishing.
15. Rysstad T, Roe Y, Haldorsen B, Svege I and Strand LI. **Responsiveness and minimal important change of the Norwegian version of the Disabilities of the Arm, Shoulder and Hand questionnaire (DASH) in patients with subacromial pain syndrome.** *BMC Musculoskelet Disord.* 2017; **18**:248. | [Article](#) | [PubMed Abstract](#) | [PubMed FullText](#)
16. Garrison C and Cook C. **Clinimetrics corner: the Global Rating of Change Score (GRoC) poorly correlates with functional measures and is not temporally stable.** *J Man Manip Ther.* 2012; **20**:178-81. | [Article](#) | [PubMed Abstract](#) | [PubMed FullText](#)
17. Guyatt GH, Townsend M, Keller JL and Singer J. **Should study subjects see their previous responses: data from a randomized control trial.** *J Clin Epidemiol.* 1989; **42**:913-20. | [Article](#) | [PubMed](#)

### Citation:

Halili A. **Patient Identified Problem (PIP) Scale, Validity, Reliability, Responsiveness, Likelihood ratio, and Minimal Clinically Important Difference.** *Phys Ther Rehabil.* 2020; **7**:4.  
<http://dx.doi.org/10.7243/2055-2386-7-4>