Original Research Paper

Reliability of Physical Examination, Diagnosis and Treatment of Low-Back Pain by Physiotherapy Practitioners in Some Selected Hospitals in Zambia

Aggrey Musanu[1], Margaret M. Mweshi[2], Moses Mwango[2]

[1]Department of Physiotherapy, Choma General Hospital, Southern Province, Choma, Zambia
[2]Department of Physiotherapy, School of Medicine, University of Zambia, Lusaka, Zambia

Email: srmmweshi@gmail.com

Abstract

Background: Low back pain (LBP) is a common disorder that has great consequences in terms of human sufferings and costs related to treatment, disability and lost productivity world over. Reliable physical examination and diagnosis are a fundamental elements in the effective clinical management of LBP. Clinicians like physiotherapists are expected to base their clinical and treatment decisions on physical examination results and diagnosis of LBP. We carried out a study to evaluate the reliability of the physical examination, diagnosis and treatment of LBP by physiotherapy practitioners in five selected referral hospitals in Zambia.
Methodology: A cross sectional study design utilising descriptive correlational techniques were employed for the study. A total of 25 physiotherapy practitioners (n=13 degree and n=12 diploma holders) were recruited for the study. The study included 100 patients aged between 18- 60 years with LBP referred to the physiotherapy department at five referral hospitals across Zambia. A standardized physical examination form was used by raters to capture the required clinical information. The reliability assessments were done using the intrarater and interrater tests. A total of 50 patients were each examined separately and consecutively by two physiotherapy practitioners (interrater) while the remaining 50 patients were examined by single therapists who repeated the examination on the same patient after three days (intrarater). Data was entered using SPSS version and descriptive statistics and Cohen's kappa statistical test was used to assess the level of agreement between the two clinical findings obtained (minimum κ: 0.61).
Results: The interrater and intrarater reliability for the assessment of LBP were; physical examination results (κ: 0.39; fair; range 0.09-0.83) and (κ: 0.62; good; range 0.06-0.94), for diagnosis (κ: 0.65; good; range -0.13-1.00) and (κ: 0.86; very good; range 0.43-1.00) and treatment (κ: 0.66; good; range 0.10-1.00) and (κ: 0.69; good; 0.23-1.00) respectively. Furthermore, the results showed that 14% and 62% of the raters scored above the minimum priori in the interrater and intrarater reliability respectively of physical examination category, while diagnosis had 52% and 92% and treatment had 56% and 60%.
Conclusion: Physical examination techniques among physiotherapy practitioners for LBP varied greatly and had inadequate reliability especially interrater relaibility. The diagnosis and treatment of LBP were considered reliable among physiotherapy practitioners from some selected hospitals in Zambia. A clinically applicable, valid and reliable classification system of physical examination for physiotherapy practitioners may be essential for future high quality research studies to be able to demonstrate the efficacy of diagnosis and physiotherapy management of LBP in Zambia.

Key words: Physiotherapy, Reliability, Physical Examination, Diagnosis, Treatment, Low Back Pain

1. Introduction

Low back pain (LBP) has great consequence in terms of human suffering, treatment costs and indirect costs related to disability and lost productivity [1]. It refers to spinal and paraspinal symptoms in the lumbosacral region [2]. It is managed by many different health care providers, among them, physiotherapists. It is reported to be a common disorder world over and approximately 60% to 80% of all adults will develop LBP sometime in their life and at any given point in time 15% to 28% of the population will have symptoms [3]. This picture is similar to that obtained in Zambia as given by

WHO Scientific Group Report [4] which estimated the point prevalence to be between 12 - 32%. Anecdotal reports show that LBP is one of the common reasons for patients seeking health care services in Zambia. It is among the top five leading causes of physiotherapy consultations by outpatients in many major clinics and hospitals representing about 20- 40% of all patients seen.

It has been demonstrated that there are several problems associated with determining a specific patho-anatomical diagnosis in many cases of LBP. As a result, the formulation of a management strategy guided by a patho-anatomically based diagnosis is not possible for a large number of the patients with LBP [5]. Traditionally, much of the evaluation of LBP has been targeted at finding rare non-mechanical spinal disorders [2]. Because most patients with LBP do not report radiating leg symptoms or other risk factors for a serious back condition, a brief physical examination is usually done in the initial phases. Unfortunately, the presence and location of soft tissue tenderness are poorly reproducible, thereby leading to physical examination disparities and consequently, unreliable physical examination results.

Physical examination and evaluation procedures have been studied extensively in people with LBP. Reliable physical examination and diagnosis are a fundamental element in the effective management of LBP. Therefore, it is imperative that physical examination findings are interpreted by clinicians with a high level of reliability for them to have clinical significance and utility [6]. Improved examination skills are well known to lead to improved diagnosis and differential diagnosis of lower extremity radiculopathy [7]. Physical therapists are expected to use physical examination (PE) procedures routinely when making clinical decisions in the management of LBP, although there is limited data available to guide the clinical decision making process. Thereby, they mostly rely on inductive reasoning, intuition, and evidence to formulate clinical decisions [8]. It is therefore not surprising that surveys of practice of physiotherapists have revealed low use of classification schemes despite evidence that treatment of patients based on subgrouping results has better outcomes than treatment based on clinical guidelines [9].

Given the reported physical examination disparities and unreliable physical examination results by physiotherapy practitioners, it cannot be disputed that valid and reliable physical examination results are very essential in the management of LBP. It is therefore important to appreciate that a clear understanding of validity and reliability in psychometrics is essential for practitioners in diverse medical settings. As Foster and Cone [10] note, "Science rests on the adequacy of its measurement. Poor measures provide a weak foundation for research and clinical endeavors". In essence, measurement pervades almost every facet of our lives and daily activities in which we measure a great variety of things [11]. Ultimately, reliable physical examination results of LBP by physiotherapy practitioners in back management is indispensable. Psychometrics is the science of assessing the measurement characteristics of scales and this process is used

to evaluate the quality of measurements [12]. Psychometric properties include different forms of reliability, validity and responsiveness [13-17]. Reliability is the degree of consistency or dependability with which an instrument could measure something [18,19]. The test-retest reliability which is a measure of temporal stability is the process of administration of an instrument to the same person at different times [18,20]. Measurement of agreement is also a reliability evaluation which is known as inter-rater reliability. The inter-rater reliability holds the consistency of performance among different raters or judges in assigning scores to the same objects or responses when two or more raters judge the performance of one group of subjects at the same point in time [15,16,18].

With the reported low usage of classification schemes in assessing patients with LBP, effective management of LBP still remains a great challenge to many physiotherapy practitioners. This can also be mainly attributed to the complex nature of LBP which makes it difficult for most of them to come up with specific and reliable diagnosis and only a small proportion (approximately 20%) of LBP cases can be attributed with certainty to a pathologic or anatomical entity [21]. According to Schäfer and colleagues [22], effective clinical management of LBP is that which allows clinicians to base treatment decisions on a reliable and valid diagnosis leading to treatment choices that result in demonstrable outcomes in terms of pain relief and functional improvement. Therefore, the aim of this study was to evaluate the reliability of the physical examination, diagnosis and treatment of LBP by physiotherapy practitioners in selected hospitals in Zambia.

2. Methodology

Upon obtaining approval from ERES Converge Institutional Review Board, permission was sought from the Hospital Directors of the five hospitals being: the University Teaching Hospital (UTH), Choma General Hospital (CGH), Mansa General Hospital (MGH), Lewanika General Hospital (LGH) and Levy Mwanawasa General Hospital (LMGH) which are spread across four different provinces of Zambia. All participants were provided with written informed consent.

A cross sectional study design utilising descriptive correlational techniques was undertaken. Each of the five study sites contributed 5 physiotherapy practitioners to obtain the number required for the study. A total of 25 physiotherapy practitioners (n=13 degree and n=12 diploma holders) of not less than two years of experience were purposively recruited and were paired according to the level of qualification. The study also conveniently recruited 100 patients with LBP aged between 18-60 years referred to the physiotherapy department at the five referral hospitals. Patients were excluded if the cause of their LBP was already known; one attributed to current pregnancy, acute fracture, osteoporosis, primary malignancy, confirmed spinal fracture within 6 months, with a history of lumbosacral spinal surgery, known infective spinal disease like TB spine, congenital spinal deformity and inflammatory systemic diseases involving the spine.

Aggrey Musanu, Margaret M. Mweshi, Moses Mwango. *Reliability of Physical Examination Results, Diagnosis and Treatment of Low-Back pain by Physiotherapy Practitioners in Some Selected Hospitals in Zambia.*

93

A standardized physical examination form adapted from Airaksinen and others [23] was used by the physiotherapy practitioners to capture the required clinical information. The assessment form (appendix 1) has five sections capturing observations, palpations, movements, neurological examinations and additional information. Diagnoses of LBP were clustered in nine classifications being; disc syndrome, nerve root entrapment, nerve root compression, spinal stenosis, zygapophsial joint, dysfunction, postural, myofascial and inconclusive [24]. With regards treatment, electrotherapy techniques assessed were infrared, short wave, hot packs, interferential and tens while others treatments assessed were exercise, massage, manipulation and lastly, a combination [24].

All physiotherapy practitioners (raters) underwent an orientation training focusing on the data collection and entry procedure to ensure that the correct methodology was followed and precise data obtained. The raters were provided with a structured patient assessment forms (appendix 1) which were used to capture patients' clinical information from the physical examination and radiograph (x-ray). These forms were used for both interrater and intrarater reliability evaluation (each physiotherapy practitioner conducted a complete patient physical examination and recorded the findings independently). The researcher identified study coordinators for each study site to oversee the procedures and ensure that the right methodology was followed. The completed patient assessment forms were collected by the study coordinator immediately after the rater concluded the patient assessment. This was done to ensure that no information regarding the patient examination was exchanged between the paired or single physiotherapy practitioners. Patients were not informed about findings of the first examination to avoid biasing the second examination.

A total of 50 patients were each examined separately and consecutively by two physiotherapy practitioners (interrater) while the remaining 50 patients were examined by single therapists who repeated the examination on the same patient after three days (intrarater). All eligible patients were simultaneously examined twice on the same day by paired raters with each rater taking about 30-40 minutes with a 10 minutes break between the examinations to allow them recuperate. The whole procedure lasted approximately 1h 20 minutes and each therapist was blinded to the other therapists' assessments as well as each patient's outcome measure. In order to maintain anonymity and avoid rater bias, no names were entered on the assessment form; the patients were identified using their OPD file numbers, while the physiotherapy practitioners and the hospitals were allocated codes.

Data was entered in the databases of SPSS version 19 and analysed using descriptive statistics for demographic characteristic and frequencies while the Cohen's kappa statistical test was used to assess the level of agreement between the two clinical findings obtained (minimum κ: 0.61). Cohen kappa ranging from -1 to +1 was interpreted as follows: values ≤ 0 as indicating no agreement, <0 .2 as poor, 0 .21–0 .40 as fair, 0 .41– 0 .60 as moderate, 0 .61–0 .80 as Good, and 0 .81–1 .00 as very good agreement.

3. Results

Demographic Characteristics

One hundred LBP patients (37 males and 63 females) were examined; fifty (n= 50) by two raters (interrater) and other 50 by single raters (intrarater). The 25 physiotherapy practitioners (10 men, 15 women) who performed the patient examinations had clinical experience of ranging from 3-17 years in managing patients with LBP, and 12 of them were diploma holders while 13 had first degrees. Table 1 shows the raters' demographic profiles.

Tables 1: Raters' Demographic Profiles

Characteristics	n (%)
Gender	
Male	*10 (40%)*
Female	*15 (60%)*
Level of Education	
Diploma	*12 (48%)*
Degree	*13 (52%)*
Institution of Employment	
University Teaching Hospital	*5 (20%)*
Levi Mwanawasa Hospital	*5 (20%)*
Lewanika General Hospital	*5 (20%)*
Choma General Hospital	*5 (20%)*
Manasa General Hospital	*5 (20%)*
Work Experience in Years	
2-6 years	*8 (32%)*
Over 6 years	*17 (68%)*

Physical Examination of Low Back Pain

The physical examination process considered five parts being: observation, palpation, movement, neurological examination and additional information. Given the main purpose of the study, the interrater and intrarater reliability evaluations were done on the physiotherapy practitioners. Table 2 presents the Kappa coefficient scores for Physical Examination of LBP. Figures 1 and 2 show the interater and intrarater reliability results of the physical examinations assessment. The results showed that 14% and 62% of the raters scored above the minimum priori in the interrater and intrarater reliability respectively. The average kappa scores being κ: 0.39 for the interrater ranging from 0.09-0.83 and κ: 0.62 for the intrarater ranging from 0.06-0.94. The frequency distribution for the data shown in figure 1 showed that most of the interrater kappa scores (κ) were falling between 0.2 and 0.6 and under a normal distribution curve (indicating a normal dispersion of data) with a mean SD of 0.394 ± 0.202. A one sample t-test statistical analysis of the kappa scores at 95% Confidence Interval (CI) gave a p value of 0.001. Nonetheless, the intrarater frequency distribution analysis shown in the figures revealed that the majority of the data was falling between 0.4 and 0.9 and was also under the distribution curve despite it being slight skewed to the right with a Mean SD of 0.62 ± 0.233. The one sample T-test statistical analysis of the kappa scores at 95% Confidence Interval (CI) provided a p value of 0.08. Physical examination techniques among physiotherapy practitioners for LBP varied greatly and had inadequate reliability especially interrater relaibility.
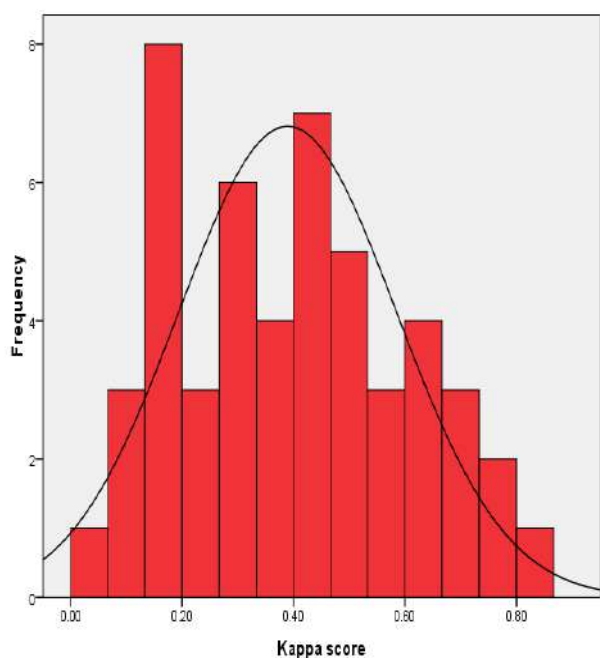

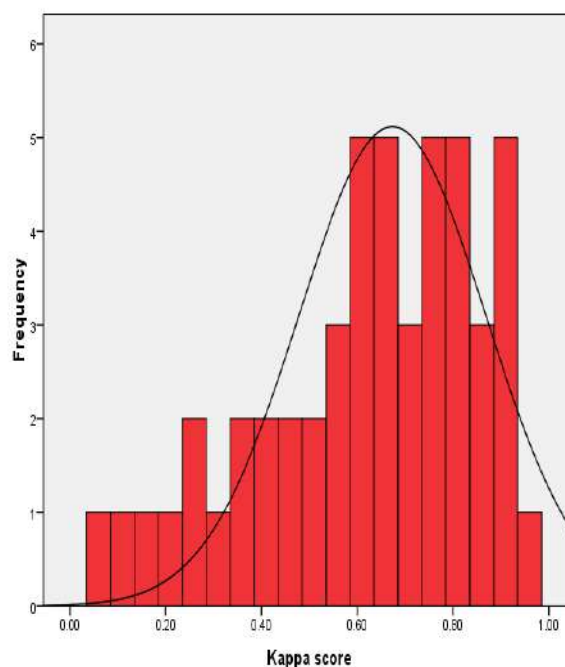*Figure 1: Interater Reliability for physical examination of LBP*

Figure 2: Intrarater Reliability for physical examination of LBP

Diagnosis of Low Back Pain

Table 3 shows the kappa coefficient scores for the diagnosis of LBP. The results showed that 52% and 92% of the raters scored above the minimum priori in the interrater and intrarater reliability respectively. The average kappa scores being κ: 0.65 for the interrater ranging from -0.13-1.00 and κ: 0.86 for the intrarater ranging from 0.43-1.00. The frequency distribution for the data tabulated in figure 3 showed that most of the interrater kappa scores were falling under the normal distribution curve with a slight shift to the right. The largest part of the data was falling between 0.4 and 1.0 giving a Mean SD of 0.671 ± 0.163 and p-value of 0.05.

Figure 4 presents the frequency distribution of the intrarater reliability (the mean, standard deviation and the distribution curve). The data (kappa score) was negatively skewed; 88% of the kappa values were falling between 0.7 and 1.0 with a Mean SD of 0.855 ± 0.146 and p-value of 0.003. About 70% of κ was falling under the distribution curve. The diagnosis of LBP were considered reliable among physiotherapy practitioners who took part in the study.
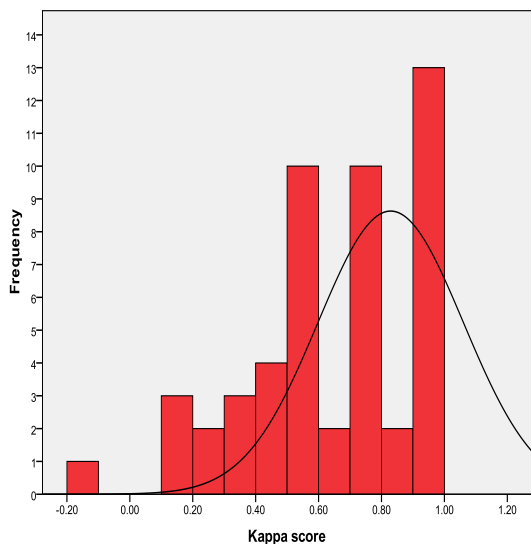
Aggrey Musanu, Margaret M. Mweshi, Moses Mwango. *Reliability of Physical Examination Results, Diagnosis and Treatment of Low-Back pain by Physiotherapy Practitioners in Some Selected Hospitals in Zambia.*

95

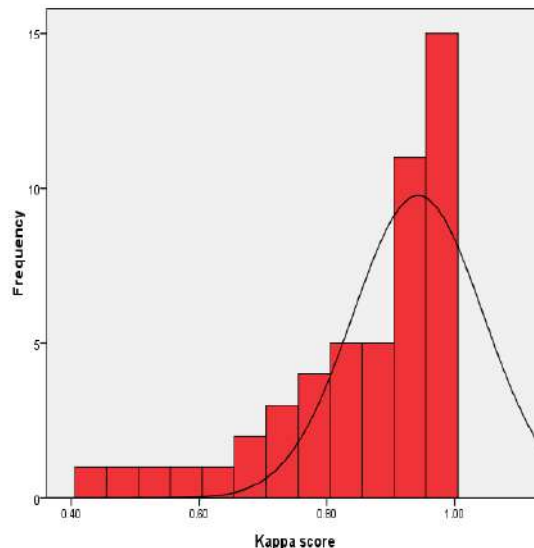*Figure 3: Interater Reliability for the diagnosis of LBP*

Figure 4: Intrarater Reliability for the diagnosis of LBP

Treatment of Low Back Pain

Table 4 presents the Kappa coefficient scores for the treatment of LBP while figures 5 and 6 show the frequency distribution for the data tabulated in the interrate and intrarater evaluation of the treatment results. The results showed that 56% and 60% of the raters scored above the minimum priori in the interrater and intrarater reliability respectively. The average kappa scores being κ: 0.66 for the interrater ranging from 0.10-1.00 and κ: 0.69 for the intrarater ranging from 0.23-1.00. Figures 5 and 6 present the frequency distribution (the mean, standard deviation and the distribution curve) for the interrater and intrarater data. It was noticed that about 50% of the interrater data (kappa score) was falling under the normal distribution curve which was negatively skewed with most of the data ranging between 0.4 and 1.0; thus giving a Mean SD of 0.662 $\pm$ 0.269. However the intrarater frequency distribution showed that most of the data (kappa score) was also falling under the normal distribution curve with the largest part falling between 0.4 and 1 with a slight shift to the right (negatively skewed) with a Mean SD of 0.69 $\pm$ 0.189 and p-value of 0.007. The treatment modalities of choice for LBP were considered reliable among physiotherapy practitioners who took part in the study
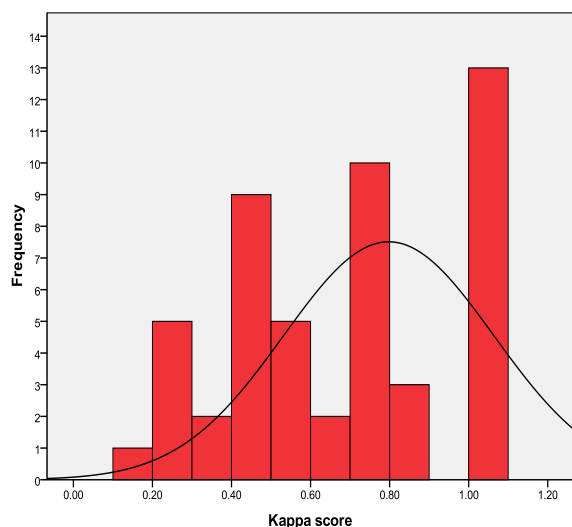

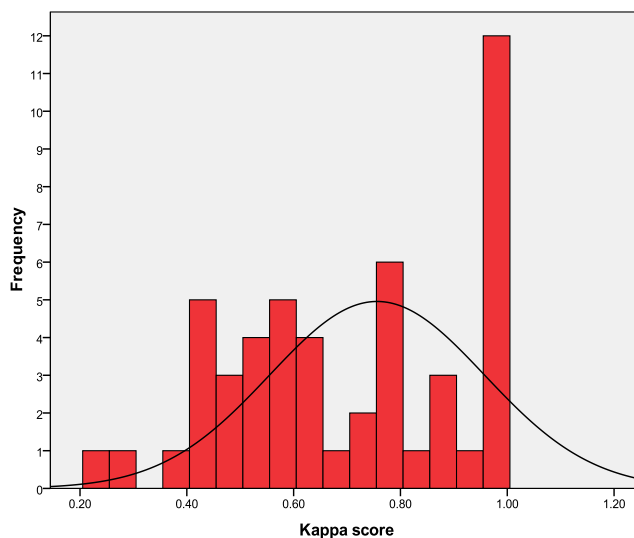*Figure 5: Interater Reliability for the treatment of LBP*

Figure 6: Intrarater Reliability for the treatment of LBP

Table 2: Kappa coefficient scores for Physical Examination of LBP

Assessment Number	INTERATER		INTRARATER	
	Kappa Statistic	Strength of Agreement	Kappa Statistic	Strength of Agreement
Assessment 1	0.6	Moderate	0.58	Moderate
Assessment 2	0.53	Moderate	0.74	Good
Assessment 3	0.44	Moderate	0.8	Good
Assessment 4	0.74	Good	0.7	Good
Assessment 5	0.62	Good	0.8	Good
Assessment 6	0.59	Moderate	0.62	Good
Assessment 7	0.7	Good	0.19	Poor
Assessment 8	0.78	Good	0.34	Fair
Assessment 9	0.54	Moderate	0.89	Very Good
Assessment 10	0.32	Fair	0.18	Poor
Assessment 11	0.31	Fair	0.65	Good
Assessment 12	0.49	Moderate	0.75	Good
Assessment 13	0.4	Fair	0.93	Very Good
Assessment 14	0.18	Poor	0.63	Good
Assessment 15	0.52	Moderate	0.84	Very Good
Assessment 16	0.2	Poor	0.74	Good
Assessment 17	0.1	Poor	0.41	Moderate
Assessment 18	0.83	Very Good	0.6	Moderate
Assessment 19	0.36	Fair	0.53	Moderate
Assessment 20	0.18	Poor	0.44	Moderate
Assessment 21	0.42	Moderate	0.68	Good
Assessment 22	0.19	Poor	0.91	Very Good
Assessment 23	0.19	Poor	0.54	Moderate
Assessment 24	0.13	Poor	0.87	Very Good
Assessment 25	0.14	Poor	0.65	Good
Assessment 26	0.46	Moderate	0.56	Moderate
Assessment 27	0.43	Moderate	0.24	Poor
Assessment 28	0.22	Fair	0.83	Very Good
Assessment 29	0.7	Good	0.42	Moderate
Assessment 30	0.53	Moderate	0.94	Very Good
Assessment 31	0.33	Fair	0.65	Good
Assessment 32	0.19	Poor	0.25	Poor
Assessment 33	0.6	Moderate	0.83	Very Good
Assessment 34	0.6	Moderate	0.29	Fair
Assessment 35	0.18	Poor	0.74	Good
Assessment 36	0.42	Moderate	0.8	Good
Assessment 37	0.68	Good	0.71	Good
Assessment 38	0.58	Moderate	0.48	Moderate
Assessment 39	0.43	Moderate	0.62	Good
Assessment 40	0.09	Poor	0.06	Poor
Assessment 41	0.3	Fair	0.38	Fair
Assessment 42	0.47	Moderate	0.89	Very Good
Assessment 43	0.34	Fair	0.72	Poor
Assessment 44	0.36	Fair	0.09	Poor
Assessment 45	0.23	Fair	0.49	Moderate
Assessment 46	0.27	Fair	0.93	Very Good
Assessment 47	0.39	Fair	0.63	Good
Assessment 48	0.12	Poor	0.84	Very Good
Assessment 49	0.15	Poor	0.74	Good
Assessment 50	0.32	Fair	0.68	Good
Number of valid kappa statistic	50		50	
Number of scores > 0.61	7 (14%)		31 (62%)	

Key: | Lowest score | Highest score |

Aggrey Musanu, Margaret M. Mweshi, Moses Mwango. *Reliability of Physical Examination Results, Diagnosis and Treatment of Low-Back pain by Physiotherapy Practitioners in Some Selected Hospitals in Zambia.*

97

Table 3: Kappa coefficient scores for the diagnosis of LBP

Assessment Number	INTERATER		INTRARATER	
	Kappa Statistic	Strength of Agreement	Kappa Statistic	Strength of Agreement
Assessment 1	0.76	Good	0.71	Good
Assessment 2	0.43	Moderate	1	Very Good
Assessment 3	1	Very Good	0.91	Very Good
Assessment 4	1	Very Good	0.92	Very Good
Assessment 5	1	Very Good	1	Very Good
Assessment 6	1	Very Good	0.68	Good
Assessment 7	0.76	Good	1	Very Good
Assessment 8	0.13	Poor	0.76	Good
Assessment 9	0.36	Fair	1	Very Good
Assessment 10	0.58	Moderate	0.81	Very Good
Assessment 11	0.42	Moderate	0.64	Good
Assessment 12	0.13	Poor	0.67	Good
Assessment 13	0.54	Moderate	0.96	Very Good
Assessment 14	0.57	Moderate	1	Very Good
Assessment 15	0.76	Good	0.54	Moderate
Assessment 16	0.76	Good	0.73	Good
Assessment 17	0.76	Good	1	Very Good
Assessment 18	0.57	Moderate	0.86	Very Good
Assessment 19	1	Very Good	0.57	Moderate
Assessment 20	1	Very Good	0.89	Very Good
Assessment 21	1	Very Good	0.76	Good
Assessment 22	0.26	Fair	0.93	Very Good
Assessment 23	0.86	Very Good	0.87	Very Good
Assessment 24	0.42	Moderate	0.86	Very Good
Assessment 25	0.76	Good	1	Very Good
Assessment 26	0.57	Moderate	0.94	Very Good
Assessment 27	1	Very Good	0.77	Good
Assessment 28	0.58	Moderate	0.91	Very Good
Assessment 29	1	Very Good	0.76	Good
Assessment 30	1	Very Good	0.92	Very Good
Assessment 31	0.58	Moderate	1	Very Good
Assessment 32	0.75	Good	0.91	Very Good
Assessment 33	1	Very Good	0.83	Very Good
Assessment 34	0.76	Good	1	Very Good
Assessment 35	0.63	Good	0.82	Very Good
Assessment 36	0.41	Moderate	0.49	Moderate
Assessment 37	1	Very Good	0.91	Very Good
Assessment 38	0.38	Fair	1	Very Good
Assessment 39	0.57	Moderate	0.83	Very Good
Assessment 40	0.23	Fair	0.71	Good
Assessment 41	0.58	Moderate	0.96	Very Good
Assessment 42	1	Very Good	1	Very Good
Assessment 43	0.76	Good	0.43	Moderate
Assessment 44	0.79	Good	0.83	Very Good
Assessment 45	0.13	Poor	0.95	Very Good
Assessment 46	0.87	Very Good	0.86	Very Good
Assessment 47	0.36	Fair	0.98	Very Good
Assessment 48	-0.13	Poor	1	Very Good
Assessment 49	0.57	Moderate	0.94	Very Good
Assessment 50	0.6	Moderate	0.93	Very Good
Number of valid kappa statistic	**50**		**50**	
Number of scores > 0.61	**26 (52%)**		**46 (92%)**	

Table 4: Kappa coefficient scores for the treatment of LBP

RELIABILTY RESULTS FOR TREATMENT				
Assessment Number	INTERRATER		INTRARATER	
	Kappa Statistic	Strength of Agreement	Kappa Statistic	Strength of Agreement
Assessment 1	0.36	Fair	1	Very Good
Assessment 2	0.26	Fair	0.45	Moderate
Assessment 3	0.44	Moderate	0.41	Moderate
Assessment 4	0.76	Good	1	Very Good
Assessment 5	0.41	Moderate	0.46	Moderate
Assessment 6	0.26	Fair	0.45	Moderate
Assessment 7	1	Very Good	0.41	Moderate
Assessment 8	0.57	Moderate	0.97	Very Good
Assessment 9	1	Very Good	0.57	Moderate
Assessment 10	0.41	Moderate	0.75	Good
Assessment 11	0.26	Fair	0.58	Moderate
Assessment 12	0.26	Fair	0.3	Fair
Assessment 13	0.1	Very Good	0.23	Fair
Assessment 14	1	Very Good	0.41	Moderate
Assessment 15	0.57	Moderate	0.41	Moderate
Assessment 16	0.76	Good	0.45	Moderate
Assessment 17	0.75	Good	0.49	Moderate
Assessment 18	1	Very Good	1	Very Good
Assessment 19	0.44	Moderate	0.76	Good
Assessment 20	0.41	Moderate	0.97	Very Good
Assessment 21	1	Very Good	0.46	Moderate
Assessment 22	1	Very Good	0.72	Good
Assessment 23	0.26	Fair	0.82	Very Good
Assessment 24	0.41	Moderate	0.57	Moderate
Assessment 25	1	Very Good	0.57	Moderate
Assessment 26	0.41	Moderate	0.63	Good
Assessment 27	0.63	Good	0.68	Good
Assessment 28	0.56	Moderate	0.76	Good
Assessment 29	0.75	Good	0.97	Very Good
Assessment 30	0.49	Moderate	0.65	Good
Assessment 31	1	Very Good	1	Very Good
Assessment 32	0.85	Very Good	0.55	Moderate
Assessment 33	1	Very Good	0.96	Very Good
Assessment 34	0.75	Good	0.61	Good
Assessment 35	0.31	Fair	1	Very Good
Assessment 36	0.44	Moderate	1	Very Good
Assessment 37	0.76	Good	0.87	Good
Assessment 38	0.81	Very Good	0.38	Fair
Assessment 39	0.74	Good	0.97	Very Good
Assessment 40	1	Very Good	0.86	Very Good
Assessment 41	0.57	Moderate	0.56	Moderate
Assessment 42	1	Very Good	0.76	Good
Assessment 43	0.89	Very Good	1	Very Good
Assessment 44	0.64	Good	0.63	Good
Assessment 45	0.76	Good	0.98	Very Good
Assessment 46	1	Very Good	0.76	Good
Assessment 47	1	Very Good	0.76	Good
Assessment 48	0.57	Moderate	0.86	Very Good
Assessment 49	0.76	Good	0.76	Good
Assessment 50	0.76	Good	0.48	Moderate
Number of valid kappa statistic	50		50	
Number of scores > 0.61	28 (56%)		28 (56%)	

The overall interrater and intrarater Cohen kappa reliability coefficients results for physical examination, diagnosis and treatment of LBP calculated at 95% Confidence Interval ranged from 0.39 to 0.86 as shown in *Table 5.*

Table 5: Kappa Reliability Coefficients for physical examination, diagnosis and treatment of LBP

Variable	Reliability Test	Mean Cohen Kappa Score (95% CI)		Standard Deviation (SD)	P-value
		Score	Interpretation		
Physical Examination	Interrater	0.39	Fair	0.202	0.001
	Intrarater	0.64	Good	0.199	0.010
Diagnosis	Interrater	0.65	Good	0.288	0.050
	Intrarater	0.86	Very Good	0.146	0.003
Treatment	Interrater	0.66	Good	0.269	0.050
	Intrarater	0.69	Good	0.232	0.007

Priori: Health Care Related Studies (K ≥ 0.61).

Aggrey Musanu, Margaret M. Mweshi, Moses Mwango. *Reliability of Physical Examination Results, Diagnosis and Treatment of Low-Back pain by Physiotherapy Practitioners in Some Selected Hospitals in Zambia.*

99

4. Discussion

This study investigated the interrater and intrarater reliability among physiotherapy practitioners in conducting physical examination, diagnosis and treatment of patients with LBP. It is essential that physical examination findings are interpreted by clinicians with a high level of reliability for them to have clinical significance and utility. In addition to having clinical significance and utility, the results must be generalizable to clinical practice [22]. Hence this study was conducted under the daily clinical conditions, with 25 (n = 25) physiotherapy practitioners examining 100 (n = 100) patients with LBP. It is with this reason that the result obtained from this study can be considered representative of everyday physiotherapy practice in across the country. An account of the study findings for the interrater and intrarater reliability of physical examination, diagnosis and treatment are discussed in this discussion.

Reliability for Physical Examination, Diagnosis and Treatment of LBP

The interrater and intrarater reliability results for physical examination, diagnosis and treatment of LBP revealed a fair to very good strength of agreement. These results can be compared with some certainty to the findings by Hicks and colleagues [25] in their study to identify lumbar segmental instability and Billis and colleagues [26] with the study looking at Interrater reliability of discriminatory examination for non-specific LBP. These results are further discussed in details under the following sub-heading:

1. Reliability for Physical Examination of LBP
2. Reliability for Diagnosis of LBP
3. Reliability for Treatment modalities of LBP.

Reliability for Physical Examination of LBP

The interrater reliability results for physical examination suggested a low level of reliability of K = 0.39 against a lower limit of 0.61, with a p-value of 0.001 (<0.05). These results denotes a statistically significant low interrater reliability of physical examination of LBP amongst physiotherapy practitioners (both diploma and degree holders). The results of this study are comparable to the findings by Kottner and colleagues [27] and Tidstrand & Horneij [28] who in their study involving interrater reliability, generated generally low mean kappa scores ranging from 0.35 to 0.62.

However the intrarater reliability results for physical examination from this study indicated good reliability within the same therapist with a mean Kappa score of (κ) = 0.62. This result further showed that the data was negatively skewed with a large standard deviation of ± 0.233 and p-value of 0.08 (> 0.05). Therefore the skewed frequency distribution coupled with a wide standard deviation (SD) and a p-value greater than 0.05 suggested rater memory bias attributed to the short period

of 3 days between the first and second assessment. According to Di Eugenio and colleagues [29], this further suggested that intrarater results for physical examination may be unreliable despite obtaining a mean kappa value of (κ) = 0.62 which close to the lower limit 0.61. This view is supported by Zenk and colleagues [30] and McHugh [31] who in their studies observed that increasing the period of time between the first and second assessment reduced the reliability levels due to decreased observer memory bias. Di Eugenio and colleagues [29], further stated that κ is affected by the skewed distributions of categories and by the degree to which raters disagree. However this study could not allocate more time between the first and second assessment because some symptoms assessed under physical examination were unstable (not fixed) hence expected to change with time. Therefore increasing the intrarater period may have resulted in a wider variation between the first and second assessment, which would have made it difficult to ascribe the variation to the test being unreliable or a change in the symptoms.

This study further suggested that some components such as observation, palpations, movements (passive and active) and neurological examination measured under physical examination were multidimensional with their own varied levels of reliability. Therefore, the reliability levels for both interrater and intrarater may have further been affected by the variability in the findings obtained under the various categories of physical examination. This observation is in agreement with the findings by Bills and colleagues [26], who in their study looking at Interrater reliability of discriminatory examination for non-specific LBP noticed variations in reliability levels under the various categories of physical examination which ranged from fair to very good agreement.

Reliability for Diagnosis of LBP

The interrater reliability results for the diagnosis of LBP indicated a high level of agreement, with a kappa value of (κ) = 0.65 against a lower limit of 0.61 and a P-value; p = 0.05. This result therefore expressed good reliability for the diagnosis of LBP made by two different physiotherapy practitioners which was also statistically significant. However this result did not measures up to the intrarater reliability result for the diagnosis of LBP which demonstrated a statistically significant high level of agreement of Kappa (κ) = 0.86 with a p-value = 0.003. Therefore the intrarater evaluation for diagnosis of LBP made by single raters on individual patients demonstrated very good reliability as compared to the one made by two different raters. The variation between the two Means for the diagnosis reliability results exhibited similar pattern to that of the interrater and Intrarater reliability physical examination results in this study. In both instances for the physical examination and diagnosis, the intrarater results exhibited higher reliability in comparison the interrater results. This observation is consistent with the findings by Liao and colleagues [32] who in their literature review study to determine the relationship between interrater and intrarater

reliability, demonstrated that intrarater results had generally higher reliability in comparison to the interrater.

However despite the similar patterns noted between physical examination and diagnosis results, physical examination recorded lower kappa scores in comparison to that of diagnosis. This difference was attributed to the limited and specific answers provided under diagnosis in contrast to the very wide categories and items which were evaluated under physical examination. This view is supported by McHugh [31] who in her study of Interrater reliability and the kappa statistic, observed that variables with fewer possible states which are sharply differentiated are likely to have high reliability levels unearthing an example on the study of survival of sepsis patients where the outcome variable had only two options; either survived or did not survive [31]. These results were also consistent with the findings by Rubinstein and colleagues [33], who in their study of diagnostic procedures for neck and low-back pain established relatively reliable diagnosis for nerve-root compression of $\kappa = 0.66$.

Reliability for Treatment Modalities of LBP

The interrater reliability results for the choice of treatment modalities for LBP showed similarity to that of the diagnosis, indicated fairly high agreement with mean kappa value of $(\kappa) = 0.66$ which was also statistically significant $(p = 0.05)$. This result expressed good reliability on the choice of treatment modalities for the one hundred (n =100) LBP patients based on the clinical examination conducted by the 25 (n = 25) dual raters. These results were consistent with the findings by Boon and colleagues [34] who in their study of the "Inter-Rater Reliability of Auditing Physical Therapists Workers' Compensation Board" generated similar range of reliability levels of $\kappa = 0.47 - 0.82$ for most of the treatment modalities which included exercises therapy, hydrotherapy, manual therapy and home programme.

These findings further demonstrated a similar pattern to that of the intrarater reliability results for the treatment of LBP in this study which also indicated a fairly high level of agreement of kappa value; $\kappa = 0.69$ with a P-value; $p = 0.007$. This result also expressed good reliability on the choice of treatment for the LBP patients made by single raters after two clinical examinations made on single patients. According to Zenk and colleagues [35] the similarity in the intrarater and interrater reliability for the choices of treatment suggested an adequate definition and understanding of the treatment modalities available for management of LBP. This view is supported by De Stefano [36] in her study of Implications of Poor Inter-Rater Reliability Outcomes Commonly Used to Diagnose Low Back Pain stated that the utility of the diagnosis-treatment model depends on the input of accurate and reliable clinical examination. She further explained that exclusive of reliable clinical examination the diagnosis cannot be used to formulate a treatment or evaluating the effects of therapeutic intervention.

5. Conclusion

Physical examination techniques among physiotherapy practitioners for LBP varied greatly and had inadequate reliability especially interrater reliability attributing this to lack of validity in the tool used for physical examination of LBP. The diagnosis and treatment of LBP were considered reliable among physiotherapy practitioners from some selected hospitals in Zambia. A clinically applicable, valid and reliable classification system of physical examination for physiotherapy practitioners may be essential for future high quality research studies to be able to demonstrate the efficacy of diagnosis and physiotherapy management of LBP in Zambia.

Acknowledgement

Conflict of interest declaration

None

References

1. Murphy D and Hurwitz E. A theoretical model for the development of a diagnosis-based clinical decision rule for the management of patients with spinal pain. *BMC Musculoskeletal Disorders 2007, 8:75:1471-2474.*

2. Atlas SJ and Deyo RA. Evaluating and managing acute low back pain in primary care setting. *J Gen Intern Med, 2001: 16: 120-131.*

3. Werneke M, Hart D, Oliver D, McGill T, Grigsby D, Ward J, Weinberg J, Oswald W and Cutrone G. Prevalence of classification methods for patients with lumbar impairments using the McKenzie syndromes, pain pattern, manipulation, and stabilization clinical prediction rules. *Journal of Manual and Manipulative Therapy. 2010. 18:4:197.*

4. Suri P, Hunter D, Katz J, Li L, Rainville J. The Burden of Musculoskeletal Conditions at the Start of the New Millennium, Report of a WHO Scientific Group, 2005.

5. Van Dillen LR, Sahrmann SA, Norton BJ et al. Reliability of Physical Examination items used for classification of patients with low back pain. *Phys Ther. 1998, 78: 979-988.*

6. May S, Littlewoods C and Bishop A. Reliability of procedures used in the physical examination of non-specific low back pain: A systematic review. *Australian Journal of Physiotherapy 2007, 52: 91–102*

7. Miller JK. Physical Assessment of lower extremity of radiculopathy and sciatica. *Journal of Chiropractic Medicine 2007: 6: 75-82.*

8. Lyle MA, Sahara M. Michael G, Sarah Z, Maura D. Relationship of Physical Examination Findings and Self-reported symptoms severity and social functions in patients with degenerative lumbar conditions. *Phys. Ther. 2005; 85: 120-133.*

Aggrey Musanu, Margaret M. Mweshi, Moses Mwango. *Reliability of Physical Examination Results, Diagnosis and Treatment of Low-Back pain by Physiotherapy Practitioners in Some Selected Hospitals in Zambia.*

101

9. Karayannis N, Jull G and Hodges P. Physiotherapy movement based classification approaches to low back pain: comparison of subgroups through review and developer/expert survey. *BMC Musculoskeletal Disorders 2012, 13:24.1471-2474.*

10. Foster, S. L., & Cone, J. D. "Validity issues in clinical assessment", *Psychological Assessment 1995, 7, 248-260.*

11. Pedhazur, E. J. and Schmelkin, L. P. Measurement, design and analysis: An integrated approach. Lawrence Erlbaum Associates Publishers: 1991; Hillsdale, New Jersey.

12. Ramaker, C., Marinus, J., Stiggelbout, A. M. and VanHilten, B. J. 2002. "Systematic evaluation of rating scales for impairment and disability in Parkinson's disease", *Movement Disorders 2002; 17, 867-876.*

13. Burton, L. J. and Mazerolle, S. M. "Survey Instrument Validity Part I: Principles of Survey Instrument Development and Validation in Athletic Training Education Research", *Athletic Training Education Journal 2011;* 6, (1), 27-35.

14. Kaun-Lin C., Ching-Lin, H., Ching-Fan, S., Fu-Chang, H. and Mei-Hui, T. 2009. "Reliability and validity of a Chinese version of the Pediatric Evaluation of Disability Inventory in Children with Cerebral Palsy, *Journal of Rehabilitation Medicine 2009;* 41(4), 273-8.

15. Domholdt, E. Rehabilitation Research. 3rd Edition, Principles and Applications: 2005; Elsevier.

16. Long, T. M. and Toscano, K. "Handbook of Pediatric Physical Therapy". 2nd edition, Mosby International Limited: 2002, Elsevier Limited.

17. McCarthy, M. L., Silberstein, C. E., Atkins, E. A., Harryman, S. E., Sponseller, P.D. and Hadley-Miller, N. A. "Comparing reliability and validity of pediatric instruments for measuring health and wellbeing of children with spastic cerebral palsy," *Developmental Medicine & Child Neurology 2002;* 44, 468–76.

18. Waltz, C. F., Strickland, O. L. and Lenz, E.R. Measurement in Nursing and Health Research. (3rd Ed): Springer Publishing Company, 2010, United States of America.

19. Cook, D. A. and Beckman, T. J. "Current concepts in validity and reliability for psychometric instruments: theory and application", *American Journal of Medicine 2006, 119: 166e7-166e16.*

20. Dekker, J., Dallmeyer, A. J. and Lankhorst, G. J. 2005. "Clinimetrics in rehabilitation medicine: current issues in developing and applying measurement instruments". *Journal of Rehabilitation Medicine 2005; 3*7, 193–201.

21. Hall H, McIntosh G, Boyle C. Effectiveness of a low back pain classification system. *Spine J 2009; 9:648–57.*

22. Schäfer and colleagues. Schäfer A, Hall T, Lüdtke K, Mallwitz J and Briffa N. Interrater reliability of a new classification scheme for patients with neural low back-related leg pain. *Journal of Manual & Manipulative Therapy 2009; 17:109–116.*

23. Airaksinen O, Brox JI, Cedrashi C, Hildebrandt J, Klaber-Moffet J, Kovacs F, et al. Working Group on Guidelines for Chronic Low Back Pain. European guidelines for the management of chronic nonspecific low back pain. *Eur Spine J 2008; 15: 192–300.*

24. 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. *BMC Musculoskeletal Disorders. 2006; 6:28.*

25. Hicks G, Fritz J, Delitto A, Mishock J. Interrater reliability of clinical examination measures for identification of lumbar segmental instability. *Arch Phys Med Rehabil 2008; 84:1858–64.*

26. Billis et al, (2012) Bills E, McCathy C, Gliatis J, Gittns M, Papandreou M and Oldham J. Inter-tester reliability of discriminatory examination items for sub-classifying non-specific low back pain. *Journal of Rehabilitation Medicine. 2012; 44: 447–453*

27. Kottner J, Audige L, Brorson S, Donner A, Gajewskie B, Hrobjartsson A, Roberts C, Shoukri M and Streiner D. Guidelines for Reporting Reliability and Agreement Studies Journal of Clinical Epidemiology 2011; 96 -106

28. Tidstrand J and Horneij E. Inter-Rater Reliability of Three Standardized Functional Tests in Patients with Low Back Pain. *Musculoskeletal Disorders 2009. 10:58 .1471-2474.*

29. Di Eugenio B, Fossati D, Haller S and Glass. Natural Language Generation for Intelligent Tutoring Systems: a case study. 2008 International Conference on Artificial Intelligence in Education, Amsterdam, Netherlands.

30. Zenk SN, Schulz AJ, Graciela et al. Inter-rater and test-retest reliability: Methods and Results for the neighbourhood observational checklist. *Health and Place 2007; (13): 452-465.*

31. McHugh M. Interrater reliability: the kappa statistic. Biochemia Medica 2012; 22 (3):276-82. National University, San Diego, California.

32. Liao S, Hunt E and Chen W. Comparison between inter-rater reliability and inter-rater agreement in performance assessment. Pub Med Singapore. 2010 Aug; 39(8):613-8.

33. Rubinstein, M and Van Tulder. A best evidence review of diagnostic procedures of neck and low back pain. *Best Prac Res Clinic Rheumatology 2008; 3: 471-482.*

34. Boon A, Quinn K, Schlichter M, Tam C and Young L. Inter-Rater Reliability of Auditing Physical Therapists Workers' Compensation Board Charts for Patients with Mechanical Low Back Pain. University of Alberta. 2010. www.physicaltherapy.ualberta.ca

35. Browder D, Childs J, Cleland J, Julie M Fritz. Effectiveness of an Extension-Oriented Treatment Approach in a Subgroup of Subjects with Low Back Pain: A Randomized Clinical Trial. *Physical Therapy. 2007; 87:1608 –1618.*

36. DeStefano L. Implications of Poor Inter-Rater Reliability Outcomes in Functional Evaluations Commonly Used to Diagnose Low Back Pain. Department of Osteopathic Manipulative Medicine, Michigan State University 2012.

APPENDIX

ASSESSMENT FORM

PHYSICAL EXAMINATION FORM

OBSERVATION

Postural assessment

Exaggerated Lordosis	Yes	No
Kyphosis	Yes	No
Sway back	Yes	No
Scoliosis	Yes	No

Gait assessment

Trendelenburg gait	Negative	Positive
Antalgic gait	Negative	Positive
Neurological gait	Negative	Positive

PALPATION

Pattern of pain

Back dominant	Yes	No
Buttock dominant	Yes	No
Leg dominant	Yes	No
Intermittent	Yes	No
Constant (Always present)	Yes	No

Pain centralization (Level of pain on the spine)

T12	Yes	No
L1	Yes	No
L2	Yes	No
L3	Yes	No
L4	Yes	No
L5	Yes	No
S1	Yes	No
S2	Yes	No

MOVEMENT

Range of motion — Normal — Reduced

Extension
Flexion
Rotation
Side flexion (Lt)
Side flexion (Rt)

Functional movement that reproduces pain

Worse in trunk flexion	Yes	No
Less pain in trunk flexion	Yes	No
No Change in flexion	Yes	No
Worse in trunk extension	Yes	No
Less pain in trunk extension	Yes	No
No effect in extension	Yes	No
Worse in walking	Yes	No
Less pain in waking	Yes	No
No effect in walking	Yes	No
Worse in sitting	Yes	No
Less pain in sitting	Yes	No
No effect in sitting	Yes	No

NEUROLOGICAL EXAMINATION

Myotomes (Muscle Power)

	Grade 0	Grade1	Grade 2	Grade 3	Grade 4	Grade 5
Trunk flexor						
Trunk extensor						
Hip flexors						
Hip extensors						
Knee flexors						

Aggrey Musanu, Margaret M. Mweshi, Moses Mwango. *Reliability of Physical Examination Results, Diagnosis and Treatment of Low-Back pain by Physiotherapy Practitioners in Some Selected Hospitals in Zambia.*

103

Knee extensors
Planter flexors
Dorsi flexors

Dermatomes (Location of symptom on the Lower limb)

	Normal	Pain	Numbness
L1			
L2			
L3			
L4			
L5			
S1			
S2			

Reflexes

Patella reflex	Normal		Exaggerated

Special tests

Straight leg raise test	Positive		Negative
Femoral nerve test	Positive		Negative
Slump test	Positive		Negative

ADDITIONAL INFORMATION (X-RAY)

Disc height (Intervertebral space) changes:	None	Mild	Moderate	Severe
Spondylothesis (vertebral displacement)	None	Mild	Moderate	Severe
Osteophytes formation	None	Posterior	Anterior	All- round

DIAGNOSIS...

TREATMENT..