6-2019

The validation of Swahili version of the international prostate symptom score questionnaire

Miten Ramesh Patel

AGA KHAN UNIVERSITY

Postgraduate Medical Education Programme

Medical College, East Africa

THE VALIDATION OF SWAHILI VERSION OF THE INTERNATIONAL PROSTATE SYMPTOM SCORE QUESTIONNAIRE

By

DR. MITEN RAMESH PATEL

A dissertation submitted in partial fulfilment of the
requirements for the degree of
Master of Medicine in General Surgery

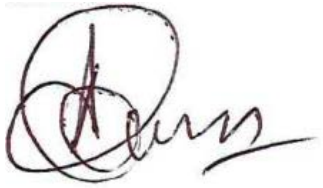
Dar es Salaam, Tanzania.

14/06/2019

DEPARTMENTAL DISSERTATIONS COMMITTEE APPROVAL

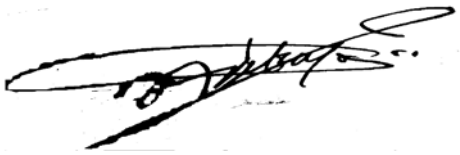
Dr Athar Ali

Chief Internal Examiner



Dr. Ali Akbar Zehri

Supervisor



Dr. Philip Adebayo

Supervisor

Dr. Athar Ali

Supervisor

AGA KHAN UNIVERSITY

Postgraduate Medical Education Programme

Medical College, East Africa

Submitted to the Board of Graduate Studies

In part fulfillment of the requirements for the degree of

Master of Medicine in General Surgery

Members of the Dissertations Standards Committee appointed to vet the dissertation of

DR. MITEN RAMESH PATEL

find it satisfactory and recommend that it be submitted for evaluation by external examiners

Chair, Dissertations Standard Committee

14th JUNE 2019

Date

DEDICATION:

To my loved ones,

Who always believed!

ABSTRACT

Background: Benign prostatic hyperplasia is a common condition encountered in men who are 50 years and above, presenting in hospitals for evaluation of lower urinary tract symptoms. The International Prostate Symptom Score is one of the questionnaires used to assess these symptoms. However, this questionnaire has not yet been validated among patients with Benign Prostate Hypertrophy in the Tanzanian setting.

Objectives: The primary objective of this study was to assess the validity of the Swahili version of the IPSS questionnaire among patients with benign prostatic hyperplasia.

Methods: We employed a cross-sectional study design in patients that presented at the Aga Khan Hospital, Dar es Salaam, Tanzania with lower urinary tract symptoms. To validate our Swahili version of the International Prostate Symptom Score questionnaire validity and reliability were calculated. The validity of the questionnaire was established with face validity and discriminant validity. The reliability of the questionnaire was established by assessing the test-retest reliability and the internal consistency of the Swahili version of International Prostate Symptom Score questionnaire. The sensitivity to change was assessed using paired T-test. Correlation coefficients were presented in tables in the relevant domain.

Results: There was excellent internal consistency observed between the Swahili International Prostate Symptom Score and the original International Prostate Symptom Score with a Cronbach's α of 0.86 and 0.919 respectively. Test-retest reliability showed high interclass correlation of 0.84. The average improvement after treatment on the Swahili International Prostate Symptom Score was 9.69 ± 6.36.

Conclusion: The Swahili International Prostate Symptom Score is reliable, valid and sensitive to change in the Tanzanian population.

LIST OF ABBREVIATIONS USED

LUTS	-	Lower Urinary Tract Symptoms
BOO	-	Bladder Outlet Obstruction
BPH	-	Benign Prostate Hyperplasia
ICIQ-MLUTS	-	International Consultation on Incontinence Questionnaire. Male Lower Urinary Tract Symptoms
DAN-PSS-1	-	Danish Prostatic Symptom Score
ICSQoL	-	International Continence Society–Benign Prostatic Hyperplasia study quality-of-life
BPH-QoL9	-	Urolife- benign prostatic hypertrophy health-related quality-of-life questionnaire
IPSS	-	International Prostate Symptom Score
AUA	-	American Urological Association
SCI	-	International Scientific Committee
WHO	-	World Health Organization
UICC	-	International Union Against Cancer
HRQoL	-	Health-related quality of life
PGWBI	-	Psychological General Well-Being Index
EQ-5D	-	EuroQol-5D
VAS	-	Visual Analogue Scale

ACKNOWLEDGEMENT

I am grateful to my supervisors, whose scholarly advice, help and constant encouragement has contributed significantly to the completion of this study.

I wish to thank my Dissertation Committee members for their critical input for my study.

I also wish to thank the management, staff, faculty members, and my fellow residents for their invaluable input and for being a great source of support to me during my study.

My gratitude to the library staff for their support.

I am grateful to my wife, Kavitha and my parents for their patience and priceless support.

I also wish to thank Jasmit Shah and Dr Ibrahim Sendagire for their invaluable input during the development of this study.

Thank you all

DECLARATION

I declare this dissertation does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any university and that to the best of my knowledge it does not contain any material previously published or written by another person except where due reference have been made in the text.

_____

(Signature of candidate)

14th JUNE 2019

Date

TABLE OF CONTENTS:

DEPARTMENTAL DISSERTATIONS COMMITTEE APPROVAL ii
DEDICATION: iv
ABSTRACT v
LIST OF ABBREVIATIONS USED vi
ACKNOWLEDGEMENT vii
DECLARATION viii
TABLE OF CONTENTS: ix
LIST OF TABLES: xi
LIST OF FIGURES: xii
1.0 INTRODUCTION AND LITERATURE REVIEW: 1
1.1 Background 1
1.2 Quality of Life 1
1.3 Health Related Quality of Life 1
1.4 Disease specific Quality of Life 2
1.5 Benign prostatic hyperplasia 2
1.6 Lower urinary tract symptoms 3
1.7 Symptom severity scores: 4
1.8 International Prostate Symptom Score 6
1.9 Validity and reliability 7
1.9.1 Validity 7
1.9.2 Reliability 9
1.10 Reliability of different translations of the IPSS Questionnaire 9
1.11 Sensitivity to change 10
1.12 Justification 11
1.13 Research questions 11
1.14 Objectives 12
1.14.1 Primary objective 12
1.14.2 Secondary objectives 12
1.15 Research hypothesis 12
2.0 METHODS: 13
2.1 Study design: 13
2.2 Study setting: 13
2.3 Study population: 13
2.3.1 Eligibility criteria: 13

2.3.2 Control Subjects: 14
2.4 Sampling procedure and recruitment: 14
2.5 Sample size: 14
2.6 Research questionnaire: 15
2.7 Data collection 15
2.8 Data management: 15
2.9 Data analysis 16
2.9.1 Validity: 16
2.9.2 Reliability 16
2.9.3 Sensitivity to change 16
2.9.4 Sensitivity and specificity 17
2.10 Ethical consideration: 17
3.0 RESULTS: 18
3.1 Reliability: 19
3.2 Sensitivity to change 21
3.3 Discriminant validity 23
4.0 DISCUSSION 25
4.1 Validity 26
4.2 Reliability 26
4.3 Sensitivity to change 27
4.4 Sensitivity and specificity 28
5.0 CONCLUSIONS 29
6.0 LIMITATIONS 30
7.0 REFERENCES: 31
8.0 APPENDICES: 37
8.1 Appendix 1: Questionnaire (English) 37
8.2 Appendix 2: Questionnaire (Swahili) 38
8.3 Appendix 3: Letter of permission form MAPI Research 40

LIST OF TABLES:

Table 1: Sociodemographic variables of patients with BPH and control subjects 18

Table 2: The distribution of responses to individual items in the sIPSS in patients with BPH and controls. 19

Table 3: Validity and reliability. 20

Table 4: Sensitivity to change; mean scores before and after treatment, mean difference between the scores, effect size and Guyatt Statistic. 21

Table 5: Swahili IPSS item characteristics based on the area under the ROC curve .23

Table 6: sIPSS item characteristics: discriminant validity.....23

Table 7: Translation comparison.....24

LIST OF FIGURES:

Figure 1: Relationship of types of validity. 8

Figure 2: Shows the ROC curve for our Swahili IPSS. 22

1.0 INTRODUCTION AND LITERATURE REVIEW:

1.1 Background

Benign prostatic hypertrophy (BPH) is a commonly encountered condition, with prevalence growing among men older than 50 years of up to 50% (1). BPH becomes clinically significant when it is associated with subjective symptoms named lower urinary tract symptoms (LUTS). The International Prostate Symptom Score (IPSS) questionnaire is a commonly used instrument in research and clinical setting to assess the severity of LUTS and guide treatment modalities as well as response to treatment. The IPSS is self-administered and only requires a few minutes for completion by patients. The original IPSS questionnaire was developed by Berry et al. in English. However, this questionnaire has not been validated in the Tanzanian setting. A validated Swahili version will help better assessment of symptom severity of LUTS and eventually guide treatment.

1.2 Quality of Life

The World Health Organization (WHO) defines health as "A state of complete physical, mental and social well-being not merely the absence of disease." It also states that assessment of health and its changes includes not only disease profile but also an estimation of one's wellbeing and this can be weighed by measuring change in Quality of Life (QoL).(2)

WHO defines Quality Of Life as "an individual's perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns."(3)

1.3 Health Related Quality of Life

There is no accepted definition of the notion of health-related quality of life (HRQoL). It goes past simply measure of health but rather it aims to measure the emotional and physical impact on the QoL.

HRQOL assessed from multiple studies concluded that patients with moderate and severe lower urinary tract symptoms reported a poor HRQoL in various domains including mental health, vitality, emotional and physical components.(4).

1.4 Disease specific Quality of Life

The disease specific QoL is not well defined, but the overall impression is ones understanding of facets of life and the array of one's activities that are impacted by the disease.

Disease specific QoL questionnaires commonly assess symptoms pertinent to the specific condition. A disease-specific measurement is important to specific diseases, both clinically and socially. It can accurately target finer aspects and domains of the disease and eventually increase the sensitivity to the individual outcomes of interest (5).

Glover et al. showed in their study that bother reflects patient's overall perception of misery of having LUTS. They showed it to reflect symptom severity, self-perception, social limitation and the bearing of LUTS. They further emphasized discomfiture and social anxiety especially relay to bother (6). It has been demonstrated in multiple studies that LUTS associated with BPH represent lower levels of overall HRQOL (4, 7, 8).

Robertson et al. also showed that severe LUTS results in a lower QoL. They showed that the impact of moderate LUTS on QoL had an impact like that of having diabetes mellitus, hypertension or cancer, while that of severe LUTS is like that of a myocardial infarction or stroke. These bearing of LUTS on QoL persisted across cultures (4).

Agrawal et al. showed in their study a strong correlation between IPSS and QOL score (9). Men with moderate to severe lower urinary tract symptoms performed poorly across important QoL dimensions. Recognition and prompt treatment of LUTS may considerably improve QoL of affected patients in these scopes (10).

A study of 200 men over 65 years old with moderate to severe LUTS were shown to have inferior QoL scores for virtually all dimensions proposed by the WHO, using the WHOQoLBref and WHOQoL-Old instruments, when compared to mildly symptomatic or asymptomatic men in similar age group (11).

1.5 Benign prostatic hyperplasia

BPH is a common condition affecting older men. BPH refers to the non-malignant growth of the prostate seen in aging men. Berry et al. showed in their study that there was no evidence of BPH in men younger than 30 years old, however the doubling time of BPH growth is 4.5 years between ages 31-50 and 10 years between ages 51-70 (1). Studies have showed high and increasing prevalence of LUTS in men older than 45 in the community and that symptoms advanced over time in a substantial proportion of men (12, 13).

In fact, prevalence of BPH based on histology, from several autopsy studies globally, showed an exponential rise of about 10% in men in their 3rd decade of life, 20% for those in their 4th decade, approached 50- 60% for men in their 6th decade, and is 80- 90% for men in their 7th and 8th decade (14, 15). Of these cases, 10.9 per 1000 men older than 80 years will eventually have significant symptoms that will need surgical intervention.

Autopsy studies conducted in Western countries showed histological prevalence of BPH in 40% of men in their 5th decade of life and approaching 90% by their 8th decade. A study by Chokkalingam et al. showed the prevalence estimates of BPH and LUTS in the west African population are comparable to those of other populations. This reflects that BPH is a major concern among men in West Africa too (16).

The pathogenesis of BPH is not completely understood. Current literature mostly highlights the role of hormonal dysregulation, growth factors that are locally released, and complex inflammatory causes (17).

Parson showed in their study that there are five categories of risk factors for BPH and LUTS. Aside from age, the other categories are modifiable lifestyle factors, inflammation, genetics, and sex steroid hormones (18).

For those with mild symptoms (IPSS score, 0–7), watchful waiting is generally advised. On the other end those patients who score moderate (8–19) or severe (20–35) on the IPSS and report significant bother from LUTS need further testing before a decision of treatment options can be made. These could include either medical or surgical treatments. Medical options include 5α-reductase inhibitors, α-adrenergic receptor blockers, or combination therapy or phytotherapeutic or herbal preparations (19). Hence its importance in guiding management.

1.6 Lower urinary tract symptoms

BPH becomes a clinical entity when it is associated with subjective symptoms. It commonly manifests as LUTS. LUTS are not specific to BPH.There are many causes of LUTS (15). LUTS can broadly be categorized into symptoms related to voiding, storage, or post micturition.

Voiding symptoms are characterized by the following symptoms; poor stream (patient's perception of slow stream as compared to his previous norm), intermittency (flow that breaks once/more when urinating), hesitancy (difficulty with initiation of stream), straining (need to use effort to initiate stream), and dysuria (discomfort during urination).

Storage symptoms are characterized by the following symptoms; urgency (a sudden desire to pass urine that can't be postponed), daytime frequency (patients feel that he voids too frequently during daytime), nocturia (the need to wake up more than once at night) and urge incontinence (involuntary leakage of urine soon after/when feels the urge to urinate).

Post-micturition symptoms are characterized by the following symptoms; sense of incomplete bladder emptying and dribble post micturition (involuntary dribble of urine after completion of urination).

BPH is a progressive disease. It may or may not be accompanied with bothersome LUTS, enlargement of the prostate, impaired urinary outflow, bladder outlet obstruction (BOO) or an increased risk of acute/recurrent/refractory urinary retention (15, 20, 21). The major aims of treatment are improvement of Quality of Life (QoL), decrease BPH-associated complications and symptomatic relief.

The estimated burden of LUTS in 2008, 2013 and 2018 were shown highest in Asia, then Europe, Africa, North America and finally South America. In the developing world these figures are projected to rise rapidly between 2008 and 2018 (Africa (30.1%), South America (20.5%) and Asia (19.7%)), but steadier in Europe (2.5).(22)

The prevalence of LUTS in 2008 and 2018 is estimated to be similar across continents, with the highest prevalence in Europe at 47.6% and 48.4%, respectively, tailed by North America at 46.3% and 47.0% respectively, Asia at 44.8% and 45.5% respectively, South America at 44.8% and 45.5% respectively and finally Africa at 43.9% and 44.2% (22).

It may be tempting to disregard BPH as an innocent disease that is an inevitable, or inconvenient result of aging, however this notion contradicts the psychological, medical and economic burden it represents (18).

1.7 Symptom severity scores:

There are many scores used to access the severity of LUTS.Some are completed by physicians. However, they have largely been replaced eg. The Boyarsky questionnaire and Madsen–Iversen questionnaire. Whereas others are self-administered; the Maine Medical Assessment Program (MMAP), Urolife- Benign Prostatic Hypertrophy Health-Related Quality-Of-Life Questionnaire (BPH-QoL9), the International Prostate Symptom Score (IPSS), Danish Prostatic Symptom Score (DAN-PSS-1), International Consultation on Incontinence Questionnaire. Male Lower Urinary Tract Symptoms (ICIQ-MLUTS), International Continence Society–Benign Prostatic Hyperplasia study QoL questionnaire (ICSmale) and

Visual Prostate Symptom Score (VPSS). Of these scores, the most commonly used score in both clinical and research is the IPSS (23-25).

The Boyarsky index was described in 1977. It mostly aimed at listing the important symptoms. These include; hesitancy, nocturia, intermittency, frequency, impaired stream, urgency, dysuria, incomplete voiding and terminal dribbling. The score was calculated by number of points (based on severity 0-3) on each of the nine questions. But its limitation was that it did not specify how the questions should be phrased. No formal validity or reliability was tested for the index until recently.

Later in 1983 Madsen-Iversen symptom score was published. This questionnaire evaluates prostate symptom severity. Like the Boyarsky index, there was no formal validity or reliability of the tool.

Maine medical assessment program introduced a self-administered symptom index in 1988 with assesses 5 symptoms (hesitancy, frequency, intermittency, dribbling, dysuria) and scored ranging between 1 to 5 for each. It had good reliability, validity and sensitivity to change how ever its use is limited by its content validity.

Danish prostatic symptom score was constructed and published in 1993 in Denmark. It accounts for 12 voiding symptoms and 3 sexual symptoms. Each of these items severity is graded and then further rated by bother factor. The severity and bother factor are multiplied and a sum of these scores is used. The voiding symptoms include nocturia, dysuria, straining, incomplete emptying, daytime frequency, hesitancy, urge incontinence, weak stream, postmicturition dribbling, stress incontinence, overflow/sleeping incontinence. The sexual function questions are pain or discomfort during ejaculation, and erection.

The advantage of this questionnaire is that each symptoms bother is individually accounted for. Thus, if a symptom does not bother a patient it gets disregarded or if a minor symptom bothers one significantly, it bears its respective weight. However, it includes some questions that are not specific to BPH. Good validation and sensitivity to change of the DAN-PSS-1 has been demonstrated by studies.(26)

The ICSmale questionnaire was published in 1996. It has twenty-two questions assessing twenty different urinary symptoms, of which nineteen also assess the bearing on that particular symptom. These include seven disease specific QoL questions and four questions regarding the sexual functioning. Its internal consistency was very high, and it was shown to be highly reliable.

Benign Prostatic Hypertrophy health-related quality-of-life questionnaire, also known as the UROLIFE questionnaire, was developed and published in 1997. Originally QoL20 contained twenty questions that are self-administered in the form of a visual analog scale; however, it was later condensed to a short questionnaire (QOL9). This includes 3 domains: wellbeing, sexual life status and BPH specific interference with activities. It is valid, reproduceable and sensitive to change.

In 2006, Chris et al. introduced the visual prostate symptom score (VPSS). This tool assesses frequency, nocturia, stream and QoL with the aid of pictogram. Multiple studies have shown positive correlation with the IPSS. The advantage of the VPSS was it can be used in patients who are illiterate or have limited education. Its limitations include its use in patients with limited vision, which is common in this age group.

1.8 International Prostate Symptom Score

The measurement committee of the American Urological Association (AUA) designed and validated a symptom index (AUA-7) for BPH(27, 28). It asks seven questions to quantify the severity of LUTS; frequency of urination, incomplete bladder emptying, straining, intermittency, urgency, nocturia and weak urine stream; all in the past month.

The patient can choose one of six answers rating from 0 to 5 (no symptom =0; at least one time = 1; less than half the time = 2; half the time =3; more than half the time =4 and always =5). The total symptom score is the sum of the responses to the seven questions, 1–7. The severity of LUTS based on the total score, can be graded as mild (0–7), moderate (8–19) and severe (20–35).

The original IPSS is internally consistent with a Cronbach alpha of 0.85 and test-retest reliability of 0.93. The total score strongly correlates with patients' global ratings of their LUTS (r = 0.78) and has been demonstrated to be sensitive to change at treatment response.

The international consensus conference in Paris 1992, recommended that future questionnairesinclude patients perceptive of the symptoms impact to their QoL. Hence the International Scientific Committee (SCI), under the patronage of the World Health Organization (WHO) and the International Union Against Cancer (UICC) recognized the AUA-7 with an addition of QoL ranging from 0-6 (delighted =0; pleased = 1; Satisfied = 2; mixed = 3; not satisfied =4; not delighted =5; and terrible = 6). This final questionnaire is what is now known as the International Prostate Symptom Score (IPSS) (29, 30). This questionnaire in English (US), is the original IPSS questionnaire. Other subsequent questionnaires were translated from this version.

This categorization guides the kind of treatment option the patient will be offered. It serves as qualitative measure of LUTS after the cause is established and can therefore be used to assess severity of disease and effectiveness of treatment. Identification of significant decline of QoL among moderate LUTS patients warrants the need of treatment (as opposed to waiting approaches), and is justification for early surgery (11, 31).

1.9 Validity and reliability

The translation underwent a linguistic validation according to the standard methodology of translation, in concordance with scheme recommended by Mapi Research Institute, 2004.

The process included the following steps:

- Forward translation,
- Backward translation and Review by clinicians,
- Cognitive debriefing and
- Proofreading to ensure the concept of the tool remains.

Just a translation using standard methodology as stated above doesn't guarantee the questionnaire would be valid or reliable in the target population. This warrants the translated questionnaire be assessed for its reliability and validity.(8)

1.9.1 Validity

Validity refers to how accurately the scale measures the concept of interest. Different types of validity broadly fall under the umbrella of construct validity. Construct validity is divided broadly into translational and criterion validity. Translational validity is further classified into face and content validity. Whereas, criterion validity is classified into concurrent, predictive, convergent and discriminant validity.

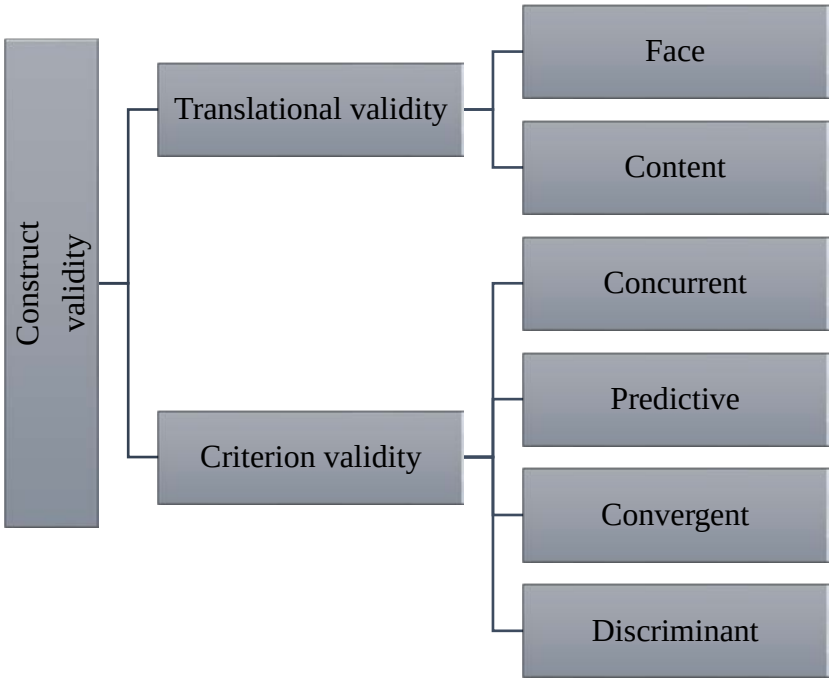


Figure 1: Relationship of types of validity.

Construct validity assesses to what level the instrument measures the idea under study (32). It can be established through a variety of ways including measuring contrasted groups, factor analysis, hypothesis testing or by MT-MM (mutitrait-multimethod) approach (33).

Translation validity assess to what level the construct has been captured into operationalization. This can be by subjective judgement ie. whether it appears to measure the construct (Face validity) or by examining its content domain. Content validity refers to qualitative measures of making sure that each item fully exploits the meaning of a domain intended by the researcher and is established using content validity ratio or content validity index (34).

Criterion validity refers to the degree of correlation between the new tool and one or more established tools (has proven validity to measure the construct of interest) (32, 34). The degree of correlation between these various tools can be measured using Spearman rank correlation coefficient. A Spearman rank correlation coefficient of ≥ 0.4 is considered acceptable (33).

Predictive validity measures the degree to which the construct predicts performance on some subsequent criterion (32, 33). Concurrent validity refers to the measure of correlation between the tool under development and the related criterion when administered simultaneously (33).

Convergent validity refers to the degree of correspondence between independent measures of the same construct. (32, 34). Independent measures in the instrument that are supposed to measure the same construct should indeed reflect high correlation coefficients. Values > 0.45

are considered adequate (32). Discriminant validity requires that scales that should not be related in reality do not have high correlation.(32, 34). Measures that assess different attributes should not have high correlations. Values < 0.45 are considered acceptable (33).

1.9.2 Reliability

Reliability is defined as how stable a measure is over a period of time (33). A reliable test should provide the similar results provided the baseline characteristics are constant (32-35).

Reliability is confirmed by calculating the test-retest reliability (33). This is done by ensuring stability of responses over time (33). The time interval can be variable ranging from a minimum of 2 weeks to 1 month has been recommended. It should be adequate to ensure the participants cannot recall their response (33). Test-retest reliability is measured using the Intra-class correlation coefficient (36). It offers insight on the degree of confidence one has, that the measure of interest reflects. It represents the correlation between the scores by estimating their linear relationship (36). An intra-class correlation coefficient value of >0.70 is considered acceptable (33).

Internal consistency estimates the degree of interrelatedness items that measure the same construct (32-35). Internal consistency can be measured with split-half reliability or Cronbach's alpha coefficient (32). The Cronbach's alpha coefficient assesses the degree to which items within the construct measure the same general concept (35, 37). A Cronbach's alpha coefficient of ≥0.9 is considered excellent and acceptable for clinical tools and ≥0.7 is acceptable for research tools (33).

1.10 Reliability of different translations of the IPSS Questionnaire

In the Spanish study by Badia et al., they got a test-retest reliability of 0.92 and Cronbach alpha of 0.79. Their results were comparable to the original IPSS (38).

In the study by Hammad et al. showed, during their validation of the Arab IPSS, achieved a the test-retest reliability of 0.88 and internal consistency coefficient 0.85 that is comparable to the original IPSS (8).

In another study, IPSS HKv2 was shown to be a reliable tool in both males and females. Their internal consistency Cronbach's alpha was 0.71 of the IPSS HKv2. The 2-week test-retest reliability was 0.7. It was comparable to other versions of the IPSS. However, they also noted lexicon difficulty in this study (39).

Quek et al. showed in thier study, of the mandarin IPSS an excellent internal consistency with a Cronbach alpha of 0.9. They also got a highly significant Test–retest correlation coefficients with Intraclass correlation coefficient of 0.91 (40).

Panahi showed in his study, internal consistency of their Persian version of IPSS was 0.7 using Cronbach's α test. Test-retest reliability was also assessed and they got an Intra class correlation coefficient of 0.87(41).

Quek et al. also did another study to validate the English IPSS in the Malaysian population, and got a high internal consistency ofCronbach alpha 0.79 and test retest ICC of 0.77(42).

1.11 Sensitivity to change

The Spanish study also assessed sensitivity to change and was higher in the Spanish version as compared to original IPSS (2.52 versus 1.44) (38). The ability of the Arabic IPSS(IPSS-Arb) to discriminate between patients with BPH and those without BPH was 0.93.(8).

1.12 Justification

There are several translations of the IPSS into different languages and these are commonly used in clinical practice and research. However, only a few translated versions have undergone a formal validation(1, 8, 38, 40, 43-45) Translating the IPSS into different languages was needed to ensure the translated instrument is abstractly comparable to the original version and acceptable to the target population. Since the IPSS is used to guide treatment modalities, it is therefore desirable that the correct categorization of the patient's symptoms is made to derive maximum benefit out of the treatment offered.

In a recent study conducted in India, Tarun et al. showed that patients who don't use English as their first language misinterpret the IPSS and that affects the treatment outcome significantly(46). Johnson et al. showed that patients who had a low level of education often misinterpret the American Urological Association Symptom Score regardless of management in a public hospital or a university practice setting. They tend to misrepresent their symptoms hence receive inappropriate treatment (44).

In Tanzania Swahili is the official language. Swahili is the language of instruction from primary education to high school, and often at higher institutions. The majority of Tanzanian population use Swahili in their day-to-day communication. Currently, a Swahili translation of IPSS is used at our institution, and probably in other hospitals. However, to the best of our knowledge based on our literature search, there is no validated Swahili version of the IPSS questionnaire. The purpose of this study, therefore, was to validate the Swahili version of the IPSS questionnaire.

1.13 Research questions

How well does the Swahili IPSS do in comparison to the original IPSS questionnaire?

1.14 Objectives

1.14.1 Primary objective

The primary objective of this study was to validate a Swahili translation of the IPSS questionnaire among patients with LUTS due to BPH attending urology clinic at the Aga Khan Hospital in Dar-es-Salaam, Tanzania

1.14.2 Secondary objectives

The secondary objectives of the present study were to assess the internal consistency and test-retest reliability of the Swahili IPSS questionnaire. We also aimed to determine the sensitivity to change of the Swahili IPSS questionnaire. We equally aimed to determine the sensitivity and specificity of the Swahili IPSS.

1.15 Research hypothesis

We expected no significance difference between the original IPSS and the final Swahili IPSS hence the following hypothesis:

H_0 : There is no significant discrepancy between the two questionnaire versions (original and Swahili)

H_1 : There is significant discrepancy between the two questionnaire versions (original and Swahili)

2.0 METHODS:

2.1 Study design:

This was an observational cross-sectional case-control study.

2.2 Study setting:

The study was conducted at Aga Khan Hospital Dar es Salaam, Tanzania, which is a tertiary level hospital and the biggest private hospital in the country with a bed capacity of 175. It has multiple sub-specialties including a well-established urology department with good operative, laparoscopy, laboratory and radiology support. The outpatient attendance at the urology clinic is over 100 patients a week. It covers patients insured by most insurance as well as those paying in cash. The Aga Khan Hospital is a teaching hospital for Aga Khan University-East Africa, Dar Es Salaam campus.

2.3 Study population:

Patients with lower urinary tract symptoms (LUTS) due to benign prostatic hyperplasia (BPH) and control subjects with suspected or confirmed urolithiasis.

2.3.1 Eligibility criteria:

2.3.1.1 Inclusion criteria:

Patients more than 50 years old attending the urology clinic between April 2018 and December 2018 with a suspected diagnosis of BPH. The diagnosis of BPH was confirmed by the urologist based on investigations or clinical criteria that included medical history and physical exam (including digital rectal examination).

2.3.1.2 Exclusion criteria:

Patients with comorbids such as uncontrolled diabetes mellitus, on diuretics, with history of previous pelvic trauma or previous surgical procedures for BPH or prostate cancer were excluded from the present study. Patients with a bladder catheter, using drugs affecting bladder function and inability to understand and answer the questionnaires were also excluded from the study.

2.3.2 Control Subjects:

Men between the age of 18 and 49-years old presenting to the Hospital with a suspected/confirmed diagnosis of urolithiasis. Patients with a diagnosis of prostatitis or BPH were not used as controls.

2.4 Sampling procedure and recruitment:

Convenience sampling was used to recruit the study participants. The patients were approached by either the primary researcher, research assistant or urology screening nurse, and were briefed about the study. If the patient agreed to participate in the study, then a formal consent was obtained by researcher or the study assistant, in Swahili or English. If agreed the patient either signed it or placed a thumbprint. Since the IPSS is a self-administered test, the patient filled it while waiting for their turn at the clinic, hence fitting in the patients' routine. They then submitted it to the researcher or assistant for collection. The translated IPSS questionnaire was administered to both the test subjects and control subjects. The IPSS questionnaire was administered again within 7 days and at 4-6 weeks to all of the original participants who received the questionnaire.

2.5 Sample size:

Many of the validation studies reviewed in the literature used a sample size between 50 and 70. A review by Anthoine published in 2014 suggests that sample size for psychometric validation may not be justified a *priori*. (47)

The formula we used to get to our sample size was (48):

$$n = 1 + \frac{2(Z_\alpha + Z_\beta)^2 k}{(\ln C_o)^2 (k-1)}$$

Where,

n= the expected sample size,

α= the probability of type I error;

β= the probability of type II error (1- power of the test);

and k= number of replicates.

We used significance level α of 0.05; power of 80%; specified correlation coefficient, $\rho 0$ of 0.9 and an expected correlation coefficient, ρ_1 of 0.95. Hence coming to a sample size of 49 participants.

An estimated attrition rate of 10% was factored in, giving us a final corrected sample size of 55 participants.

A 1:1 ratio of case to controls was used, hence we got 55 control participants, however due to scarcity of controls we modified the ratio to 2:1 and used 32 controls in our analysis.

2.6 Research questionnaire:

The translation process of the questionnaire was guided by the method approved by MAPI research. The original IPSS was translated to Swahili by a professional translator. This was followed by two independent backward translations by health care professionals. The original and the two back-translated questionnaires were compared. We found some discrepancy on 1 item, the urgency question. A discussion with the translator was moderated and the item adjusted. Similar discrepancies were noted in a study by Salman et al. however their participants were unable to understand the response option, which required them to modify theirs responces to a 6-point Likert-Scale.(49) The final adjusted questionnaire was then piloted with 10 participants, and interviewed on whether they understood the questionnaire, and whether any word or phrase was unclear. All 10 participants found the questionnaire clear and understood it well.

2.7 Data collection

Data were collected including both socio-demographics and clinical parameters. The completed questionnaire was handed to the principal investigator or the assistant. The forms were inspected for completeness of the collected data by either the principle investigator or research assistant, if complete they were finally stored under lock and key.

2.8 Data management:

Patient names were not entered on the questionnaires to ensure privacy. Each participant was allocated a unique code which was corresponded to a separate questionnaire which had all the socio-demographic characteristics of the participants. The data from the questionnaires were entered in MS Excel (2016), cleaned and finally transferred to the IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp. for further analysis.

2.9 Data analysis

2.9.1 Validity:

Several psychometric measurements were used to assess validity of the IPSS questionnaire. For translation validity, the following were assessed;

-*Face validity* – is the degree to which an instrument appears to be understandable and relevant for the targeted population [15]. Face validity was assessed during one of the steps in translating the questionnaire. Feedback from health workers working in urology clinic or theatre was incorporated regarding the effectiveness of the translated questions.

-*Discriminate validity* is the capability of the instrument to discriminate between constructs that are theoretically different. It is calculated based on scores as predicted between groups (the expectation being to score low and high in a particular trait). It was assesses the extent to which a measure differs. Discriminate validity is shown by significant differences in mean scores across independent samples.(40)

Construct validity by Spearman correlation coefficient of the IPSS with the QoL question 8.

2.9.2 Reliability

- *Reliability;* using the Test-retest method. Test-retest reliability was assessed by administering the Swahili IPSS to the same participants at initial visit and again within 7 days. The degree of correlation among the two scores, and between the individual items, indicates the stability of the instrument. Test-retest reliability was assessed using Intraclass Correlation Coefficient (ICC)

-*Internal consistency* which measures how well individual items on a construct fit together conceptually, was assessed using Cronbach's alpha coefficient.

2.9.3 Sensitivity to change

The sensitivity to change of our Swahili version of the IPSS was evaluated by comparing the mean scores at initial presentation and at 4-6 weeks after treatment (either surgical or medical) using a paired t -test. The sensitivity to change of the Swahili IPSS and its individual constructs were analyzed by calculating the difference between the scores per and post treatment and dividing it by the mean standard deviation of the scores pretreatment (effect size) (50). The Guyatt statistic was also calculated by taking the average differences in the Swahili IPSS scores pre and post-treatment and by dividing it with mean SD of the control group scores.(51)

2.9.4 Sensitivity and specificity

The sensitivity and specificity of the sIPSS were evaluated by calculating the area under the ROC curve. Sensitivity and specificity were assessed based on a cutoff value of 7.5.

2.10 Ethical consideration:

Ethical approval was obtained from Ethics and Research Committee of The Aga Khan University. Formal permission has already been obtained from the owners of the IPSS questionnaire to translate the questionnaire into Swahili language. Written permission was obtained from the Aga Khan Hospital. A written consent was obtained from all study participants prior to administering the questionnaire also explaining them that all information obtained was confidential.

3.0 RESULTS:

A total of 85 respondents took part in our validation study. The Swahili IPSS was administered to the BPH group (n = 53 mean age: 59.64 ± 7.96 years) and to the control group (n = 32; mean age: 33.21 ± 6.87 years) with a $p<0.001$. However, we had not matched our cases to controls. Table I, list the sociodemographic distribution of BPH patients and control subjects that were enrolled in the present study. Twenty-nine (54.7%) subjects had moderate urinary symptoms, 10 (18.9%) had severe LUTS while 14 (26.4%) scored mild LUTS on the sIPSS questionnaire. One patient (0.018%) had comorbids that were controlled as per the treating physician's assessment. The lowest score on the BPH group was 1, graded by only 1 patient (0.018%) while the highest score achieved was 33, graded by 1 patient (0.018%). In the control group,87.5% (n=28) scored less than 7, 9.4% (n=3) scored moderate LUTS while 3.1% (n=1) scored severe LUTS at a value of 31. The lowest score, 0 was graded by 12.5% (n=4) of the control subjects on the sIPSS. Neither cases nor control graded the maximum score in the present study. The distribution of responses to individual items on the sIPSS questionnaire for both BPH patients and control subjects is described in Table 2.

Table 1: Sociodemographic variables of patients with BPH and control subjects

		Case (n = 53)	Control (n = 32)	
Age		59.64 ± 7.96	33.21 ± 6.87	p < 0.001
IPSS	0 to 7	14 (26.4%)	28 (87.5%)	p < 0.001
	8 to 21	29 (54.7%)	3 (9.4%)	
	22 to 35	10 (18.9%)	1 (3.1%)	
Level of Education	None	2(3.8)	1(3.1)	p = 0.926
	Primary	4(7.5)	3(9.4)	
	Secondary	13(24.5)	6(18.8)	
	Tertiary	34(64.2)	22(68.8)	

IPSS: International Prostate Symptom Score

Table 2: The distribution of responses to individual items in the sIPSS in patients with BPH and controls.

Questions	Scores	0	1	2	3	4	5
		n (%)	n (%)	n (%)	n (%)	n (%)	n (%)
1-Emptying	Case	13 (24.5)	12 (22.6)	7 (13.2)	11 (20.8)	5 (9.4)	5 (9.4)
	Control	23 (71.9)	6 (18.8)	2 (6.2)	0 (0)	0 (0)	1 (3.1)
2-Frequency	Case	10 (18.9)	13 (24.5)	10 (18.9)	11 (20.8)	6 (11.3)	3 (5.7)
	Control	20 (62.5)	8 (25)	2 (6.2)	1 (3.1)	1 (3.1)	0 (0)
3-Intermittency	Case	17 (32.1)	11 (20.8)	4 (7.5)	7 (13.2)	6 (11.3)	8 (15.1)
	Control	27 (84.4)	4 (12.5)	1 (3.1)	0 (0)	0 (0)	0 (0)
4-Urgency	Case	21 (39.6)	6 (11.3)	9 (17)	10 (18.9)	2 (3.8)	5 (9.4)
	Control	29 (90.6)	2 (6.2)	0 (0)	0 (0)	0 (0)	1 (3.1)
5-Weak Stream	Case	12 (22.6)	4 (7.5)	14 (26.4)	10 (18.9)	7 (13.2)	6 (11.3)
	Control	24 (75)	3 (9.4)	4 (12.5)	0 (0)	0 (0)	1 (3.1)
6-Hesitancy	Case	20 (38.5)	7 (13.5)	10 (19.2)	7 (13.5)	3 (5.8)	5 (9.6)
	Control	26 (81.2)	2 (6.2)	2 (6.2)	1 (3.1)	0 (0)	1 (3.1)
7-Nocturia	Case	2 (3.8)	14 (26.4)	14 (20.8)	15 (28.3)	7 (13.2)	4 (7.5)
	Control	9 (28.1)	13 (40.6)	13 (15.6)	1 (3.1)	1 (3.1)	3 (9.4)

n=frequency
Responses for questions 1-6: - 0= Not at All,1=Less than 1 in 5 Times.2=Less than Half the Time,3=Half the Time, 4=More than Half the Time, 5=Almost Always
Responses for question 7: - 0=None, 1=1 Time, 2=2 Times, 3=3 Times, 4=4 Times, 5=5 Times

3.1 Reliability:

We assessed Test–retest reliability in both the patients with BPH (n=53) and the control group (n=32), and it was found that the sIPSS had an ICC of greater than 0.846 ($P < 0.001$) in both groups (Table 3). The ICC ranged between 0.846 to 1 for both the groups. The QoL question had the lowest ICC of 0.846 in the control group whereas the nocturia question scored lowest at 0.855 in the cases group.

Table 3 lists test–retest mean scores for each item of the Swahili-IPSS. It also lists the ICC and internal consistency (Cronbach alpha) results for both groups. Overall both the patients with BPH and control groups scored a high internal consistency and Inter Class Correlation. At test-retest, there were no items that were statistically significant in either groups.

The Cronbach alpha of all eight questions ranged from 0.919 to 1 in both cases and controls. The lowest Cronbach alpha was seen in the QoL question in controls and nocturia item in cases with a Cronbach alpha coefficient of 0.919.

Table 3: Validity and reliability.

	ICC*			Internal Consistency		Mean Test Score		SD		Mean Retest Score		SD		Mean Difference		SD	
	Overall	Control	Case	Control	Case	Control	Case	Control	Case	Control	Case	Control	Case	Control**	Case**	Control	Case
Incomplete emptying	0.944	1.000	0.914	1.000	0.954	0.470	2.060	1.016	1.669	0.470	1.960	1.016	1.640	0.000	0.100	0.000	0.687
Frequency	1.000	1.000	1.000	1.000	1.000	0.590	1.980	0.979	1.487	0.590	1.980	0.979	1.487	0.000	0.000	0.000	0.000
Intermittency	1.000	1.000	1.000	1.000	1.000	0.190	1.960	0.472	1.870	0.190	1.960	0.471	1.870	0.000	0.000	0.000	0.000
Urgency	1.000	1.000	1.000	1.000	1.000	0.220	1.640	0.906	1.677	0.220	1.640	0.906	1.677	0.000	0.000	0.000	0.000
Weak Stream	1.000	1.000	1.000	1.000	1.000	0.500	2.260	1.078	1.643	0.500	2.260	1.078	1.643	0.000	0.000	0.000	0.000
Hesitancy	1.000	1.000	1.000	1.000	1.000	0.440	1.630	1.105	1.675	0.440	1.630	1.105	1.681	0.000	0.000	0.000	0.000
Nocturia	0.921	0.986	0.855	0.993	0.919	1.410	2.530	1.456	1.324	1.410	2.430	1.500	1.323	0.000	0.100	0.254	0.714
QoL	0.916	0.846	0.944	0.919	0.997	2.440	3.700	2.747	1.804	2.060	3.740	2.639	1.799	0.380	-0.040	1.476	0.192

*Control group, n= 32; Case group, n = 53. QoL; Quality of Life, ICC; intraclass correlation efficient. SD; Standard Deviation. *P < 0.001 for all ICC. ** t-test for paired comparisons are not significant.*

3.2 Sensitivity to change

The sensitivity to change of the sIPSS was assessed after 1 month in BPH patients who underwent treatment. Mean age was 59.64 ± 7.96 years (range: 51 – 80 years). Fifty-two patients out of 53 had completed the questionnaire again post-treatment (Table 4). Table 4 lists scores before and after treatment, mean differences, the Guyatt statistic and effect size index for each of the items. The mean scores pre-treatment on the sIPSS for our BPH group was 14.09 ± 8.37 and the mean scores post- treatment was 6.21 ± 4.89 ($P < 0.0001$). This shows us an average improvement after treatment on the sIPSS of 7.88 ± 5.77. The effect size index of the sIPSS and sensitivity to change were both high. The effect size index for the sensitivity to change was also compared by the Guyatt statistic. A high value of Guyatt statistic showed that the pre and post-treatment scores were different. The Guyatt statistic is unlike the effect size index in that the Guyatt statistic is calculated based on the mean differences between scores pre and post-treatment and dividing these by the baseline SD of the control group.(40) Using a cutoff score of 7.5, a sensitivity of 73.6% and specificity of 67.9% was calculated.

Table 4: Sensitivity to change; mean scores before and after treatment, mean difference between the scores, effect size and Guyatt Statistic.

		Before Treatment		After Treatment					
		Mean	SD	Mean	SD	Mean Differences*	SD	ESI	Guyatt Statistic
1	Incomplete emptying	2.02	1.66	1.08	1.05	0.94	1.38	0.57	0.93
2	Frequency	1.96	1.5	0.81	0.89	1.15	1.27	0.77	1.17
3	Intermittency	1.94	1.88	0.92	1.19	1.02	1.09	0.54	2.17
4	Urgency	1.58	1.63	0.63	0.79	0.94	1.21	0.58	1.04
5	Weak Stream	2.27	1.66	0.92	0.95	1.35	1.3	0.81	1.25
6	Hesitancy	1.63	1.68	0.65	0.91	0.98	1.35	0.58	0.89
7	Nocturia	2.5	1.32	1.19	0.95	1.31	1.04	0.99	0.9
	IPSS	14.09	8.37	6.21	4.89	7.88	5.77	0.94	1.35
8	QoL	3.67	1.81	1.46	1.09	2.21	1.66	1.22	0.8

*Sensitivity to change (n = 52), SD; Standard Deviation, ESI; effect size index, ESI = mean difference/SD before treatment, Guyatt statistic = mean difference/SD of control group, IPSS; International Prostate Symptom Score, QoL; quality of life. *P < 0.0001.*

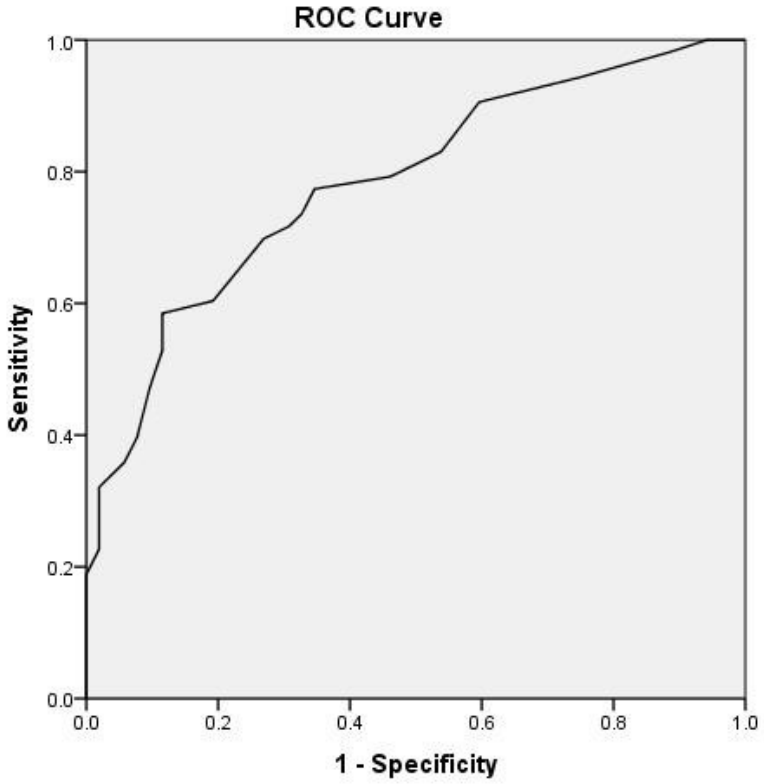


Figure 2: Shows the ROC curve for our Swahili IPSS.

The calculated area under the curve was 0.783. The ability of the individual items of sIPSS to discriminate between control subjects and those with BPH was assessed by the area under the ROC curve for individual items, as shown in Table 5. We calculated a range between 0.642–0.839, these indicates a high discriminatory power. The mean area under the ROC curve (SEM, 95% confidence interval) for the Swahili IPSS was 0.783 (0.04, 0.70–0.87). This would imply that the sIPSS would appropriately identify a randomly selected BPH patient with LUTS and a randomly selected control participant 78.3 percent of the time. The Swahili IPSS also showed a high correlation with the Swahili QoL item (Spearman rank correlation coefficient, 0.72; $p < 0.001$).

Table 5:Swahili IPSS item characteristics based on the area under the ROC curve

	Cases	AUC	Std Error	P Value
1	Incomplete emptying	**0.6420**	**0.0550**	**0.0130**
2	Frequency	**0.7260**	**0.0490**	**< 0.0001**
3	Intermittency	**0.6520**	**0.0540**	**0.0070**
4	Urgency	**0.6530**	**0.0550**	**0.0070**
5	Weak Stream	**0.7340**	**0.0510**	**< 0.0001**
6	Hesitancy	**0.6610**	**0.0540**	**0.0050**
7	Nocturia	**0.7660**	**0.0470**	**< 0.0001**
	IPSS	**0.7790**	**0.0450**	**< 0.0001**
8	QoL	**0.8390**	**0.0410**	**< 0.0001**

The area under the ROC curve was based on the cases before and after the treatment.

3.3 Discriminant validity

The discriminant validity of the sIPSS was assessed. The difference between scores in both groups for all items of the sIPSS were highly significant. The greatest change was seen in the total score of the sIPSS($p<0.001$).

Table 6: sIPSS item characteristics: discriminant validity

		Mean Test Score		SD		Mean Difference	SEM	Lower	Higher	P Value
		Control	Case	Control	Case					
1	Incomplete emptying	0.470	2.06	1.016	1.669	1.590	0.291	1.009	2.167	<0.001
2	Frequency	0.590	1.980	0.979	1.487	1.390	0.268	0.855	1.920	<0.001
3	Intermittency	0.190	1.960	0.471	1.870	1.770	0.270	1.235	2.315	<0.001
4	Urgency	0.220	1.640	0.906	1.677	1.420	0.281	0.865	1.981	<0.001
5	Weak Stream	0.500	2.260	1.078	1.643	1.760	0.295	1.177	2.353	<0.001
6	Hesitancy	0.440	1.64	1.105	1.675	1.200	0.302	0.622	1.823	<0.001
7	Nocturia	1.410	2.530	1.456	1.324	1.120	0.315	0.492	1.752	<0.001
	IPSS	3.810	14.090	5.820	8.370	10.280	1.680	6.920	13.630	<0.001
8	QoL	2.440	3.700	2.747	1.804	1.260	0.545	0.164	2.357	<0.001

The original IPSS that was developed and validated in the US had a test–retest reliability of 0.92, internal consistency of 0.86, and sensitive to change with an ESI of 1.44. This is comparable to the sIPSS in the present study, suggesting that the sIPSS is a valid, a sensitive and a reliable tool as the original IPSS.

Table 7: Translation comparison

	Original IPSS	sIPSS
Test–retest	0.920	0.840
Internal consistency	0.860	0.919
Sensitivity to change	1.440	0.940
Discriminant validity	$P < 0.0001$	$P < 0.001$
Correlation with item 8	0.770	0.720
Discriminatory power	0.850 ± 0.030	0.783 ± 0.040

4.0 DISCUSSION

The IPSS questionnaire was originally developed by American Urology Association in 1992. It was named the American Urological Association symptom score (AUA-7). Later the SCI, WHO and the UICC recognized the AUA-7 as the IPSS questionnaire after addition of QoL question. The final questionnaire consists of seven lower urinary tract symptoms and an eighth QoL questionnaire. The IPSS is a self-administered questionnaire that way it avoids any bias that would arise if administer by physician or health worker administration. This is because the IPSS is intended to pick out and grade the patient's own perception of their LUTS.(52) The IPSS questionnaire has been used in both the clinical setting and research in Africa, however it has been limited grossly owing to a major drawback of illiteracy, in this population.(53) During our literature search we did not come across any studies that validated the questionnaire in their local language in Africa. The original IPSS has been developed in different languages based on their population.(38-41, 54-56). These studies showed their versions had similar validity and reliability to the original IPSS.

The mean age of BPH patients was 59.64 (± 7.96) and that our control group was 33.21(± 6.87). these distributions was similar to that of the Arab study at 32.6 (±8.1)(8), Persian study at 61.5 (± 8.3)(41) and IPSS HKv2 had 63.2 (11.5). however, the IPSS HKv2 included both men and women.(39) The Mandarin study had a higher age group for their cases at 70.64 (± 8.51).(40)

The level of education of both the cases and controls were similar in the present study with a p value of 0.926. Two participants, one form each had missing information about their level of education. We assumed they had attained the lowest rank of education i.e. none. Abiola et al. showed in their study patients with at least secondary education were able to answer and understand the IPSS questionnaire without assistance.(52) Most of our patients 88.3%(n=75) had at least secondary level of education. However, none of the patients had asked for assistance during grading their score. This was possibly due to a higher level of education in our group of patients.

Most of the studies have used different ways to conclude their validity. This is probably because there is no concrete classification in validity studies and many of the terms are used interchangeably. However, most studies did similar analysis of their results to come to the same conclusions. This emphasizes the need for clear definitions and classifications to get clear standards in validation studies. Reliability is rather clearly defined and better understood in these studies.

4.1 Validity

All items in the present study had a good discriminate validity as evident by the high differences in mean score with $p<0.001$ between the test and control group. Thus confirming that the sIPSS was able to discriminate well between those who had LUTS and the control subjects. The mandarin study by Quek et al., also had similar findings with $p<0.001$ for the difference in mean score in both groups, hence confirming their validity too.(40) As expected Hammad et al. also found their IPSS-Arb had higher score on the cases then controls with a $p<0.001$. (8) However, one must keep in mind that these finding are based on comparing with controls that were not age matched. This does add the advantage of our validation being comparable to the validation of the original IPSS. But our controls, and those of other studies, were far from ideal. In order to exclude patients who would have LUTS or BPH in the control group, it inevitably narrows our options to a younger population.(27, 38) This inturn carries with it the limitations of the original IPSS.(38)

The Spearman rank-correlation coefficient between items 1 to 7 and the QoL of the sIPSS was good at 0.72($P<0.001$). This was close to that of the original IPSS at 0.77. Not many studies have assessed correlation. The IPSS-Sp has a similar correlation at 0.72 with the QoL question. However, they also did assess correlation with Psychological General Well-Being Index (PGWBI), EuroQol-5D (EQ-5D) and visual analogue scale (VAS). They found that their IPSS-Sp had better correlations with psychologic domains as compared to physical Domains.(38) The IPSS-Arb too had a high correlation coefficient of 0.82 with QoL item, but the study did not assess correlation with other tools.(8) Quek et al. did not assess correlation coefficient in their study.(40) Correlation coefficients greater than 0.6 reflects good correlation, 0.31 to 0.6 reflect adequate correlation, whereas below 0.3 reflect a poor correlation. The above finding confirms that our sIPSS is a valid tool and comparable to the original IPSS.

4.2 Reliability

We performed test-retest reliability and internal consistency analysis to assess reliability. Test–retest reliability was assessed in 53 participants with BPH. In the present study we found the sIPSS had an ICC of 0.846 to 1. None of the item had a statistically significant ICC. One possible reason for this high ICC was due to the short retest interval. In the present study we used an interval of less than 1 week since it fits into the normal routine and it would be ethically wrong to withhold treatment for 2 weeks. We did not expect patient symptoms to have changed before retest as treatment was initiated around the retest time, and treatment response is gradual

with maximum response at about 1 month. These findings were like that of the Mandarin study that had an excellent ICC of 0.98. (40) The Spanish study equally had an excellent ICC of 0.87.(38) The same was the case for the IPSS-Arb where, Hammad et al. got an excellent ICC of 0.88.(8) Panahi et al., attained an excellent ICC of 0.78 in their study of validation of a Persian version of the IPSS. A more recent validation by Salman et al. of the Urdu version of IPSS had an ICC of 0.92. However they validated their questionnaire in both male and females and had to modify the responses to the Likert-scale so their questions were better understood.(49)

An ICC above 0.75 is considered to have excellent test-retest reliability, whereas that between 0.4 and 0.75 as good and that of less than 0.4 as week correlation.(33) In view of this data we can conclude the sIPSS is demonstrated to be a reliable questionnaire with excellent test-retest reliability.

Cronbach alpha coefficient was calculated for internal consistency. A Cronbach's alpha coefficient of greater than or equal to 0.9 is considered excellent and acceptable for clinical tools,while that of 0.7 is acceptable for research tools (33).The Cronbach alpha of the eight questions of sIPSS ranged from 0.919 to 1. This was somewhat better than the original IPSS which had an ICC of 0.86. Similar finding were obtained by the Mandarin study that had an alpha coefficient of 0.97. (40) The Spanish study too had an excellent alpha coefficient of 0.79 though lower than that of the original IPSS it is still acceptable.(38) The IPSS-Arb had an excellent coefficient alpha of 0.85.(8) Panahi et al., too got an acceptable Cronbach alpha at 0.7 in their study of validation of a Persian version of the IPSS. Salman et al. had an alpha Coefficient of 0.72 for their Urdu version of IPSS though less, it is acceptable.

In view of these findings we can conclude that the sIPSS is reliable with excellent internal consistency and having a Cronbach alpha of 0.919 it is acceptable in both clinical and research settings.

4.3 Sensitivity to change

We assessed the sensitivity to change for the sIPSS 1-month post-treatment in BPH group. The mean pre-treatment sIPSS overall score was 14.09 ± 8.37 and the mean post-treatment score was 6.21 ± 4.89 ($P < 0.0001$). The present study showed an average improvement of 7.88 ± 5.77 after treatment on the sIPSS. The sIPSS had an effect size index of 0.94 and a high sensitivity to change. The effect size index for the sensitivity for change were compared using the Guyatt statistic. This high Guyatt value showed that the pre and post-treatment scores were

different. (40) It was lower than that of the original IPSS that had an ESI of 1.44. The Mandarin study had a somewhat better ESI at 1.66 whereas the IPSS-Sp had 2.52. One possible reason for that would be, in the Spanish study they cases had an overall higher pre-intervention score. In the present study we also demonstrated a mean improvement of 7.88 ('better') in the sIPSS was comprehended as less of an improvement in LUTS than the mean difference of 10.28 ('much better'). These findings, like those from other studies show that the change in sIPSS scores pre and post-treatment reflect meaningful clinical change.

4.4 Sensitivity and specificity

The individual sIPSS items had a high discriminant validity between patients who have BPH and control subjects. We assessed the discriminatory power by calculating the area under the ROC curve for each item and got AUC ranging between 0.642–0.839. The mean area under the ROC curve for our Swahili IPSS was 0.783 ± 0.040. This indicates the sIPSS has a high discriminatory power. This was comparable the original IPSS (0.850 ± 0.030),(27) and lower than the IPSS-Sp than had ROC of 0.50 ± 0.020. (38) Hammad et al. got a ROC of 0.93 ± 0.09 for their IPSS- Arb.(8) These findings suggest that the sIPSS has a high discriminant validity.

5.0 CONCLUSIONS

The findings of the present study show that the psychometric properties of the Swahili IPSS (sIPSS) are similar to those of the original American IPSS and other translations too. This was evident by excellent internal consistency observed between the Swahili IPSS with a Cronbach's α ranging between 0.919 and 1.000. Test-retest reliability was high ICC of 0.846 to 1.000. The average improvement after treatment on the sIPSS was 9.69 ± 6.36. Overall, we got a high effect size index and sensitivity to change. In conclusion our Swahili version of the IPSS proved to be a reliable and a valid tool that was sensitive to clinical change in assessing LUTS in men. Our finding suggests the score obtained from sIPSS are comparable to those of the original IPSS. Based on the present study findings, we recommend it can be used in both clinical and research settings in the Tanzanian population.

6.0 LIMITATIONS

Our test-retest interval was less than the standard 2 weeks to 1 month; however this was unavoidable as we cannot delay treatment due to research purposes.

Our controls were not age matched.

7.0 REFERENCES:

1. Berry SJ, Coffey DS, Walsh PC, Ewing LL. The development of human benign prostatic hyperplasia with age. J Urol. 1984;132(3):474-9.
2. WHOQOL: Measuring Quality of Life [Available from: https://www.who.int/healthinfo/survey/whoqol-qualityoflife/en/.
3. The World Health Organization Quality of Life assessment (WHOQOL): position paper from the World Health Organization. Soc Sci Med. 1995;41(10):1403-9.
4. Robertson C, Link CL, Onel E, Mazzetta C, Keech M, Hobbs R, et al. The impact of lower urinary tract symptoms and comorbidities on quality of life: the BACH and UREPIK studies. BJU Int. 2007;99(2):347-54.
5. Coyne KS, Wein AJ, Tubaro A, Sexton CC, Thompson CL, Kopp ZS, et al. The burden of lower urinary tract symptoms: evaluating the effect of LUTS on health-related quality of life, anxiety and depression: EpiLUTS. BJU Int. 2009;103 Suppl 3:4-11.
6. Glover L, Gannon K, McLoughlin J, Emberton M. Men's experiences of having lower urinary tract symptoms: factors relating to bother. BJU Int. 2004;94(4):563-7.
7. Vela-Navarrete R, Alfaro V, Badiella LL, Fernandez-Hernando N. Age-stratified analysis of I-PSS and QoL values in spanish patients with symptoms potentially related to BPH. Eur Urol. 2000;38(2):199-207.
8. Hammad FT, Kaya MA. Development and validation of an Arabic version of the International Prostate Symptom Score. BJU Int. 2010;105(10):1434-8.
9. Agrawal CS, Chalise PR, Bhandari BB. Correlation of prostate volume with international prostate symptom score and quality of life in men with benign prostatic hyperplasia. Nepal Med Coll J. 2008;10(2):104-7.
10. Welch G, Weinger K, Barry MJ. Quality-of-life impact of lower urinary tract symptom severity: results from the Health Professionals Follow-up Study. Urology. 2002;59(2):245-50.
11. Griebling TL. Re: Elderly men's quality of life and lower urinary tract symptoms: an intricate relationship. J Urol. 2012;188(3):894-5.

12. Maserejian NN, Chen S, Chiu GR, Wager CG, Kupelian V, Araujo AB, et al. Incidence of lower urinary tract symptoms in a population-based study of men and women. Urology. 2013;82(3):560-4.

13. Egan KB. The Epidemiology of Benign Prostatic Hyperplasia Associated with Lower Urinary Tract Symptoms: Prevalence and Incident Rates. Urol Clin North Am. 2016;43(3):289-97.

14. Bushman W. Etiology, epidemiology, and natural history of benign prostatic hyperplasia. Urol Clin North Am. 2009;36(4):403-15, v.

15. Roehrborn CG. Benign prostatic hyperplasia: an overview. Rev Urol. 2005;7 Suppl 9:S3-S14.

16. Chokkalingam AP, Yeboah ED, Demarzo A, Netto G, Yu K, Biritwum RB, et al. Prevalence of BPH and lower urinary tract symptoms in West Africans. Prostate Cancer Prostatic Dis. 2012;15(2):170-6.

17. Dobrek L, Thor PJ. Benign prostatic hyperplasia - progress in pathophysiology and management. Pol Merkur Lekarski. 2015;39(233):263-70.

18. Parsons JK. Benign Prostatic Hyperplasia and Male Lower Urinary Tract Symptoms: Epidemiology and Risk Factors. Curr Bladder Dysfunct Rep. 2010;5(4):212-8.

19. Foo KT. Current assessment and proposed staging of patients with benign prostatic hyperplasia. Ann Acad Med Singapore. 1995;24(4):648-51.

20. Rosen R, Altwein J, Boyle P, Kirby RS, Lukacs B, Meuleman E, et al. Lower urinary tract symptoms and male sexual dysfunction: the multinational survey of the aging male (MSAM-7). Eur Urol. 2003;44(6):637-49.

21. Girman CJ. Population-based studies of the epidemiology of benign prostatic hyperplasia. Br J Urol. 1998;82 Suppl 1:34-43.

22. Irwin DE, Kopp ZS, Agatep B, Milsom I, Abrams P. Worldwide prevalence estimates of lower urinary tract symptoms, overactive bladder, urinary incontinence and bladder outlet obstruction. BJU Int. 2011;108(7):1132-8.

23. Fitzpatrick JM. The natural history of benign prostatic hyperplasia. BJU Int. 2006;97 Suppl 2:3-6; discussion 21-2.

24. Burnett AL, Wein AJ. Benign prostatic hyperplasia in primary care: what you need to know. J Urol. 2006;175(3 Pt 2):S19-24.

25. Medina JJ, Parra RO, Moore RG. Benign prostatic hyperplasia (the aging prostate). Med Clin North Am. 1999;83(5):1213-29.

26. Hansen BJ, Flyger H, Brasso K, Schou J, Nordling J, Thorup Andersen J, et al. Validation of the self-administered Danish Prostatic Symptom Score (DAN-PSS-1) system for use in benign prostatic hyperplasia. Br J Urol. 1995;76(4):451-8.

27. Barry MJ, Fowler FJ, Jr., O'Leary MP, Bruskewitz RC, Holtgrewe HL, Mebust WK, et al. The American Urological Association symptom index for benign prostatic hyperplasia. The Measurement Committee of the American Urological Association. J Urol. 1992;148(5):1549-57; discussion 64.

28. Liu CC, Wang CJ, Huang SP, Chou YH, Wu WJ, Huang CH. Relationships between American Urological Association symptom index, prostate volume, and disease-specific quality of life question in patients with benign prostatic hyperplasia. Kaohsiung J Med Sci. 2004;20(6):273-8.

29. Rhodes PR, Krogh RH, Bruskewitz RC. Impact of drug therapy on benign prostatic hyperplasia-specific quality of life. Urology. 1999;53(6):1090-8.

30. O'Leary M P. Validity of the "bother score" in the evaluation and treatment of symptomatic benign prostatic hyperplasia. Rev Urol. 2005;7(1):1-10.

31. Wasson JH, Reda DJ, Bruskewitz RC, Elinson J, Keller AM, Henderson WG. A comparison of transurethral surgery with watchful waiting for moderate symptoms of benign prostatic hyperplasia. The Veterans Affairs Cooperative Study Group on Transurethral Resection of the Prostate. N Engl J Med. 1995;332(2):75-9.

32. Trochim WM. The Research Methods Knowledge Base Cincinnati, OH.: Atomic Dog Publishing; 2006 [Available from: www.socialresearchmethods.net/kb/.

33. DeVon HA, Block ME, Moyle-Wright P, Ernst DM, Hayden SJ, Lazzara DJ, et al. A psychometric toolbox for testing validity and reliability. J Nurs Scholarsh. 2007;39(2):155-64.

34. Drost EA. validity and reliability in social science research(online). Education Research Perspectives. 2011;38(1):105-23.

35. Health-Related Quality of Life Research | ISOQOL - International Society for Quality of Life Research 2015 [Available from: http://www.isoqol.org/about-isoqol/what-is-health-related-quality-of-life-research.

36. Vaz S, Falkmer T, Passmore AE, Parsons R, Andreou P. The case for using the repeatability coefficient when calculating test-retest reliability. PLoS One. 2013;8(9):e73990.

37. Tavakol M, Dennick R. Making sense of Cronbach's alpha. International Journal of Medical Education. 2011;2:53-5.

38. Badia X, Garcia-Losa M, Dal-Re R, Carballido J, Serra M. Validation of a harmonized Spanish version of the IPSS: evidence of equivalence with the original American scale. International Prostate Symptom Score. Urology. 1998;52(4):614-20.

39. Choi EP, Lam CL, Chin WY. Validation of the International Prostate Symptom Score in Chinese males and females with lower urinary tract symptoms. Health Qual Life Outcomes. 2014;12:1.

40. Quek KF, Chua CB, Razack AH, Low WY, Loh CS. Construction of the Mandarin version of the International Prostate Symptom Score inventory in assessing lower urinary tract symptoms in a Malaysian population. Int J Urol. 2005;12(1):39-45.

41. Panahi A. Validity and Reliability of Persian Version of International Prostate Symptom Score. Galen Medical Journal 2013(2):18-21.

42. Quek KF, Low WY, Razack AH, Loh CS. Reliability and validity of the International Prostate Symptom Score in a Malaysian population. BJU Int. 2001;88(1):21-5.

43. Homma Y, Tsukamoto T, Yasuda K, Ozono S, Yoshida M, Yamaguchi T. [Evaluation of psychometric properties of Japanese version of international prostate symptom score and BPH impact index]. Nihon Hinyokika Gakkai Zasshi. 2003;94(5):560-9.

44. Quek KF, Low WY, Razack AH, Sin Loh C, Chua CB. Reliability and validity of the Malay version of the International Prostate Symptom Score in the Malaysian population. J Urol. 2002;167(3):1359-62.

45. Sagnier PP, Richard F, Botto H, Teillac P, Dreyfus JP, Boyle P. [Adaptation and validation in the French language of the International Score of Symptoms of Benign Prostatic Hypertrophy]. Prog Urol. 1994;4(4):532-8; discussion 9-40.

46. Jindal T, Sinha RK, Mukherjee S, Mandal SN, Karmakar D. Misinterpretation of the international prostate symptom score questionnaire by Indian patients. Indian J Urol. 2014;30(3):252-5.

47. Anthoine E, Moret L, Regnault A, Sebille V, Hardouin JB. Sample size used to validate a scale: a review of publications on newly-developed patient reported outcomes measures. Health Qual Life Outcomes. 2014;12:176.

48. Bujang MA. A simplified guide to determination of sample size requirements for estimating the value of intraclass correlation coefficient: a review. Archives of Orofacial Sciences The Journal of the School of Dental Sciences, USM. 2017;12(1): 1-11.

49. Salman M, Khan AH, Sulaiman SAS, Hughes J, Khan JH, Hussain K. The Modified Urdu version of International Prostate Symptom Score: A psychometric validation study. Turk J Urol. 2018;44(4):335-40.

50. J C. Statistical Power Analysis for the Behavioural Sciences. New York, Academic Press. 1977:75–105.

51. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. J Chronic Dis. 1987;40(2):171-8.

52. Abiola OO. Use and ease of self-administered International Prostate Symptoms Score (IPSS) and Visual Prostate Symptoms Score (VPSS) questionnaires for the assessment of lower urinary tract symptoms in Nigerian men. African Journal of Urology. 2016;22(2):121-6.

53. Ogwuche EI, Dakum NK, Amu CO, Dung ED, Udeh E, Ramyil VM. Problems with administration of international prostate symptom score in a developing community. Ann Afr Med. 2013;12(3):171-3.

54. Homma Y, Tsukamoto T, Yasuda K, Ozono S, Yoshida M, Shinji M. [Linguistic validation of Japanese version of International Prostate Symptom Score and BPH impact index]. Nihon Hinyokika Gakkai Zasshi. 2002;93(6):669-80.

55. Badia X, Garcia-Losa M, Dal-Re R. Ten-language translation and harmonization of the International Prostate Symptom Score: developing a methodology for multinational clinical trials. Eur Urol. 1997;31(2):129-40.

56. Akilov FA. External validation and reliability estimation of the Uzbek and Russian version of the International Prostate Symptom Score (IPSS) questionnaire. экспериментальная и клиническая урология 2012;2:63.

8.0 APPENDICES:

8.1 Appendix 1: Questionnaire (English)

In the past month: About	Not at All	Less than 1 in 5 Times	Less than Half the Time	Half the Time	More than Half the Time	Almost Always	Your score
1. Incomplete Emptying How often have you had the sensation of not emptying your bladder?	**0**	**1**	**2**	**3**	**4**	**5**	
2. Frequency How often have you had to urinate less than every two hours?	**0**	**1**	**2**	**3**	**4**	**5**	
3. Intermittency How often have you found you stopped and started again several times when you urinated?	**0**	**1**	**2**	**3**	**4**	**5**	
4. Urgency How often have you found it difficult to postpone urination?	**0**	**1**	**2**	**3**	**4**	**5**	
5. Weak Stream How often have you had a weak urinary stream?	**0**	**1**	**2**	**3**	**4**	**5**	
6. Straining How often have you had to strain to start urination?	**0**	**1**	**2**	**3**	**4**	**5**	
	None	1 Time	2 Times	3 Times	4 Times	5 Times	
7. Nocturia How many times did you typically get up at night to urinate?	**0**	**1**	**2**	**3**	**4**	**5**	
Total I-PSS Score							
Quality of Life Due to Urinary Symptoms	Delighted	Pleased	Mostly Satisfied	Mixed	Mostly Dissatisfied	Unhappy	Terrible
If you were to spend the rest of your life with your urinary condition just the way it is now, how would you feel about that?	**0**	**1**	**2**	**3**	**4**	**5**	**6**

8.2 Appendix 2: Questionnaire (Swahili)

DALILI ZA TEZI DUME KIMATAIFA KWA FUMBO LA ISHIRINI (I-PSS)							
	Sio kabisa	Chini ya mara 1 kwa mara 5	Chini ya nusu ya muda	Takriban nusu ya muda	Zaidi ya nusu ya muda	Takriban kila mara	
1. Katika kipindi cha mwezi uliopita au takriban, ni mara ngapi hukupata hisia za kutoa mkojo wote baada ya kukojoa?	0	1	2	3	4	5	
2. Katika kipindi cha mwezi uliopita au takriban, ni mara ngapi ulirudia kukojoa kabla ya kitambo cha saa mbili baada kumaliza kukojoa?	0	1	2	3	4	5	
3. Katika kipindi cha mwezi uliopita au takriban,ni mara ngapi ulijikuta unasita kisha kuanza tena kukojoa?	0	1	2	3	4	5	
4. Katika kipindi cha mwezi uliopita au takriban, ni mara ngapi ulihisi ugumu kuairisha kukojoa?	0	1	2	3	4	5	
5. Katika kipindi cha mwezi uliopita au takriban, ni mara ngapi ulikuwa na mkojo dhaifu?	0	1	2	3	4	5	
6. Katika kipindi cha mwezi uliopita au takriban, ni mara ngapi ulijilazimisha au kutumia nguvu kukojoa?	0	1	2	3	4	5	
	Bila	Mara 1	Mara 2	Mara 3	Mara 4	Mara 5 au zaidi	
7. Katika kipindi cha mwezi uliopita, ni mara ngapi uliamka kukojoa toka uingie kitandani usiku mpaka kuamka asubuhi?	0	1	2	3	4	5	

UBORA WA MAISHA UNAOSABABISHWA NA DALILI ZA MKOJO							
	Kufurahishwa	Kupendezwa	Kuridhika Zaidi	Mchanganyiko (Kuridhika na kutokuridhika)	Kutoridhika Kabisa	Kutofurahishwa	Mbaya sana
8. Kama ungeliweza kuishi na tatizo lako la mkojo kwa maisha yako yote, utajisikiaje?	0	1	2	3	4	5	6

8.3 Appendix 3: Letter of permission form MAPI Research.

Patel Miten
Agan Khan Hospital
PGME
Ocean Road
P.O. Box 2289
Dar-Es-Salaam
Tanzania

Date: 10/19/2017

Re: Letter to certify Patel Miten rights to use and translate the I-PSS

Dear Patel Miten,

This letter is to certify that Patel Miten, Tanzania, has signed a User Agreement and a Translation Agreement with Mapi Research Trust, dated 20/06/2017, in order to gain access to use and translate the I-PSS.

The author and copyright owner, Dr Barry MJ has granted Mapi Research Trust the official exclusive right to distribute the I-PSS Questionnaire, acting on his behalf.

Mapi Research Trust hereby confirms that Patel Miten can translate the original US I-PSS into Swahili and then use the Swahili translation in the context described in the User Agreement.

Sonia Bothorel,
Operations Director

Mapi Research Trust • 27 rue de la Villette • 69003 Lyon • France
Tel: +33 (0) 4 72 13 66 66 • http://mapi-trust.org/ • www.mapigroup.com
https://eprovide.mapi-trust.org/