

Use of the International Prostate Symptoms Score (IPSS) in Chinese male patients with benign prostatic hyperplasia

Authors: Carlos King-ho Wong¹, Edmond Pui-hang Choi², Steve Wai-hee Chan³, James Hok-leung Tsu⁴, Chi-wai Fan⁵, Peggy Sau-kwan Chu⁶, Fu-keung Cheung⁷, Wai-Kit Ma⁴, Ida Soo-fun Mah⁸, Sidney Kam-hung Yip⁹, Simon See-ming Hou¹⁰, Hing-Shing So¹¹, Cindy, Lo-kuen Lam¹

¹ Department of Family Medicine and Primary Care, The University of Hong Kong

² School of Nursing, The University of Hong Kong

³ Department of surgery, Hong Kong Sanatorium & Hospital

⁴ Department of surgery, Queen Mary Hospital

⁵ Department of surgery, Pamela Youde Nethersole Eastern Hospital

⁶ Department of surgery, Tuen Mun Hospital

⁷ Department of surgery, Princess Margaret Hospital

⁸ Hong Kong Urology Clinic

⁹ Pedder Clinic

¹⁰ Department of surgery, Princess of Wales Hospital

¹¹ Department of surgery, United Christian Hospital

Correspondence Author and person to whom reprint requests should be addressed:

Carlos King Ho Wong, PhD, MPhil, BSc

Department of Family Medicine and Primary Care, The University of Hong Kong

Address: 3/F, Ap Lei Chau Clinic, 161 Ap Lei Chau Main Street, Ap Lei Chau, Hong Kong

Contact: +852-25185688 (tel); +852-28147475 (fax) carlosho@hku.hk (email)

Keywords: IPSS; benign prostatic hyperplasia; HRQOL; Chinese; Validity; Reliability

Running Title: IPSS in Chinese male with BPH.

Conflict of interest statement: This study has been funded by Hong Kong Urological Association Council. The funder had no role in study design, data collection and analysis of the manuscript. The authors declare that they have no conflict of interest.

Informed consent: Informed consent was obtained from all individual participants included in the study.

Competing interest: None declared.

Abstract

Purpose: To test the psychometric properties of the International Prostate Symptom Score Hong Kong Chinese version 2 (IPSS) in Chinese male patients with BPH under secondary care.

Methods: A prospective longitudinal study was done by interviewing subjects at baseline, at 2-week after baseline for assessing test-retest reliability and at 26-week after baseline for assessing responsiveness. All subjects were interviewed to complete a structured questionnaire including IPSS, Short Form-12 Health Survey Version 2 (SF-12v2) and Depression Anxiety Stress Scale (DASS).

Results: The IPSS HRQOL score had weak correlations with SF-12v2 summary and DASS domain scores. For reliability analysis, Cronbach's alpha coefficient was 0.90 for the seven symptom-related items. The intra-class correlation coefficients of the IPSS total symptom score and HRQOL score were 0.90 and 0.86, respectively. For sensitivity, statistically significant differences were detected between the subjects with BPH and those without for IPSS total symptom score (effect size = 0.68) but not the IPSS HRQOL score. The areas under ROC curves for the IPSS total symptom and HRQOL scores were 0.67 and 0.60, respectively.

Conclusions: The IPSS was valid, reliable instrument in Chinese patients with BPH. The IPSS total symptom score, but not the HRQOL score, is sensitive in differentiating subgroups.

Manuscript Text

Introduction

Benign prostatic hyperplasia (BPH) is a common urological condition among aging males. Although many men with a history of BPH and even anatomically enlarged prostates due to this condition have no symptoms, more than 50% of men in their 60s to as many as 90% of octogenarians present with lower urinary tract symptoms (LUTS)[1]. An epidemiological study in Hong Kong suggested that moderate-to-severe LUTS is common in elderly Chinese males while the prevalence rate of mild and moderate-to-severe LUTS was 56% and 38.2[2], respectively. Furthermore, a population-based study in Canada, Germany, Italy, Sweden, and the United Kingdom found that the prevalence of storage LUTS was 51.3%, voiding LUTS was 25.7%, and post-micturition LUTS was 16.9% in men [3]. Although LUTS is not a life threatening disease, those suffering from LUTS, especially for those with moderate and severe symptoms, have a poorer health-related quality of life (HRQOL) [4-6].

In addition to objective measures such as conducting an uroflowmetry, the severity of LUTS is assessed by a self-reported questionnaire. The International Prostate Symptom Score (IPSS) is one of the commonly-used questionnaires to assess the severity of LUTS and LUTS-specific HRQOL in male patients. The questionnaire was first developed by the American Urological Association [7]. The questionnaire is reliable, valid, sensitive and responsive [8]. The World Health Organization has subsequently suggested the IPSS is an important instrument in the assessment of BPH [9]. The role of the IPSS measurement for LUTS and BPH is well established, attributable in part to the evidence linking the severity of LUTS with hormonal status[10-12] and metabolic syndrome components[13, 14].

The IPSS was translated into Traditional Chinese by the Hong Kong Urological Association (HKUA) in 1995. The content validity and psychometric properties of the Hong Kong Chinese version of IPSS were subsequently reviewed by Choi et al [15]. The study by Choi et al found that some translations of the original Hong Kong Chinese version of the IPSS were not equivalent to the original English version. Modifications were made, resulting in the

development of the IPSS Hong Kong Chinese version 2[15]. The content validity, construct validity, reliability, sensitivity and responsiveness were evaluated and confirmed in Chinese primary care patients [15, 16]. Since the sample of patients taken for the study were Chinese primary care patients, the results might not apply to patients under secondary care, the disease profile of whom may differ. The lack of knowledge of the psychometric properties of the IPSS instrument in patients under secondary care necessitated the present study.

Aims and Objectives

The aim of this study is to test the psychometric properties of the IPSS instrument in Chinese male patients with BPH under secondary care. The objectives are the following: to assess the validity of the construct of the IPSS, to assess the reliability of the IPSS, to assess the sensitivity of the IPSS, and to assess the responsiveness of the IPSS.

Methods

Study design

This is a prospective longitudinal study. Subjects were interviewed at the baseline, 2 weeks following the baseline to assess the test-retest reliability, and 26 weeks after the baseline to assess responsiveness.

Subjects and sampling

A group of patients with BPH and a group of control subjects without BPH from the urology specialist out-patient clinics of three hospitals (Queen Mary Hospital, Tuen Mun Hospital and Princess Margaret Hospital) were recruited between June 2014 and December 2014. Patients diagnosed and documented with BPH by their physicians were eligible to join the study. Patients without a documented diagnosis of lower urinary tract symptoms and benign prostatic hyperplasia were eligible to join the control group of the study. Patients aged < 18 years, those who could not understand Cantonese, or were too ill to give consent, had cognitive impairment or dementia, or refused to join the study were excluded from both groups.

Procedure

Trained field workers approached eligible patients in three urology specialist out-patient clinics (one clinic per hospital) to invite them to participate in this observational study. The aims, objectives, procedures and nature of the study were explained before obtaining the written consent of the subjects. Subjects who gave consent were asked to provide their telephone number and were subsequently contacted by trained interviewers (from the Social Science Research Centre, the University of Hong Kong) who blindly recruited the subjects by telephone within two weeks (the baseline), and conducted blind follow-ups 2 weeks after the baseline interviews and at the 26-week mark after the baseline interviews. Telephone interviews were used to enhance response and completion rates as interviews could be conducted at a more convenient time for the respondents and reduces service disruption [17]. To maintain the quality of the data, all study

instruments were read verbatim in a standardized way. Subjects were contacted using a computer-aided telephone interviewing system by the interviewers between 10:00 am and 10:30 pm on weekdays up to a maximum of five attempts.

Study instruments

The IPSS (Hong Kong Chinese version 2)

The IPSS (Hong Kong Chinese version 2) contains seven questions regarding symptoms (incomplete bladder emptying, frequency of urination, intermittency, urgency, weak urine stream, straining and nocturia) and one LUTS-specific HRQOL question. Higher scores indicate poorer HRQOL and more severe LUTS. The psychometric properties including the validity, reliability, sensitivity and responsiveness of the IPSS have been confirmed in Chinese primary care patients [15, 16]. The total symptom score ranges from 0 to 35. A total symptom score less than or equal to 7 is mild, 8-19 represents moderate symptoms, and 20-35 represents severe symptoms. The IPSS Voiding sub-score (IPSS-V, the sum of scores from incomplete bladder emptying, intermittency, weak stream and straining) and IPSS Storage sub-score (IPSS-S, sum of scores from frequency, urgency and nocturia) are computed to calculate the IPSS voiding-to-storage sub-score ratio (IPSS V/S). Responses to the HRQOL question range from 0 to 6.

The Short Form-12 Health Survey Version 2 (SF-12v2)

The SF-12v2 is a generic HRQOL measure that covers eight domains namely physical functioning (PF), role limitation due to physical problems (RP), bodily pain (BP), general health (GH), vitality (VT), social functioning (SF), role limitation due to emotional problems (RE) and mental health (MH) [18]. The domain scores are calculated by the sum of the relevant item scores and their transformation into a representative range from 0 to 100 [19]. The SF-12v2 can be summarized into physical and mental component summary scores (PCS and MCS) with higher scores indicating a better HRQOL. The population norm of the SF-12 v2 has been established for Chinese adults in Hong Kong [20, 21].

Depression Anxiety Stress Scale 21 (DASS-21)

DASS-21 contains 3 scales to measure depression, anxiety and stress. Each scale consists of 7 questions. The multiplied scores can be categorized as normal, mild, moderate, severe and extremely severe as defined in the DASS manual. Several studies have been conducted to show the validity and reliability of the scale [22, 23]. The scale has been applied in both clinical and non-clinical settings [24, 25]. The correlations between the scale and the quality of life question of the IPSS were assessed.

Socio-demographic information

Socio-demographic information including age, education, monthly household income, and marital status was asked.

Clinical information

The parameters of urodynamic variables including the maximum urinary flow rate (Q_{max}), post-void residual urine and voided volume were retrieved from paper or electronic medical records.

Sample size calculation

The sample size for comparing the mean difference of the IPSS between the baseline and the 26-week follow-up in patients with BPH was estimated. A sample size of 130 patients in total was needed to have a 0.3 effect size, 80% power and 95% confidence interval with the a 30% attrition rate using a paired t-test. The sample size for the control group was estimated. A sample size of 85 patients in the control group was needed to have 0.85 required area under the curve (AUC) of the receiver operating characteristic (ROC) with 0.03 standard error [8, 26].

Data Analysis

Descriptive statistics, including the mean, standard deviation, and percentage of floor and ceiling scores were calculated. 15% was used as the threshold for a significant floor or ceiling effect [27]. The internal construct validity of the IPSS was assessed by examining the corrected item-total scale correlation using the corrected item-total scale correlation scores of ≥ 0.4 to identify an adequate correlation [28]. The convergent validity of the IPSS HRQOL and total symptom

scores were assessed using Pearson's correlation test against the SF-12v2 and DASS-21. It was hypothesized that the IPSS HRQOL score would have statistically significant correlations with SF-12v2 MCS and the PCS scores and DASS-21 scores, respectively. The convergent validity of the IPSS total symptom score was assessed using Pearson's correlation test against the Qmax, post-void residual urine and voided volume, respectively. It was hypothesized that the IPSS HRQOL score would have statistically significant correlations with Qmax, post-void residual urine and voided volume, respectively. The internal consistency of the IPSS was assessed using Cronbach's alpha using cut-off scores of ≥ 0.7 to indicate adequate internal consistency [29]. Test-retest reliability was assessed by examining the intra-class correlation (ICC) coefficient. $ICC \geq 0.7$ was used to indicate good reproducibility [27]. The sensitivity of the IPSS was determined by performing known group comparisons of the mean total symptom and HRQOL scores by conducting an independent t-test and by calculating the area under the receiver operating characteristic curve (AUC). It was hypothesized that (i) subjects with BPH would have higher symptom scores and poorer health-related quality of life than control subjects and (ii) the IPSS would be able to determine and differentiate patients with LUTS and those without LUTS, with an AUC of >0.7 . To assess the responsiveness of the IPSS, the change in the mean scores of the IPSS between the baseline and the 26-week mark interviews were examined conducting a paired t-test. The IPSS score differences between the baseline and 26-week interviews were evaluated by standardized effect size (SES) [30], and the standardized response mean (SRM) [31]. The effect size statistics can provide a clear interpretation of the magnitude of the change. All statistical analyses were conducting using SPSS version 23.0. A P-value of less than 0.05 was considered to be of statistical significance.

Ethics

The study protocol was approved by the institutional review board of The University of Hong Kong and the Hong Kong Hospital Authority for the Hong Kong West Cluster (reference: UW 14-058), Kowloon West Cluster (reference: KW/EX-14-032 (72-08)) and the New Territories West Cluster (reference: NTWC/CREC/1259/14).

Running Title: IPSS in Chinese male with BPH

All procedures of the study involving human participants were performed in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Written consent was obtained from all individual participants involved in the study.

Results

Baseline sociodemographic information

A total of 297 subjects were invited to participate in this observational study. From this total, 60 subjects refused to give consent to participate in the study, whereas the remaining 237 subjects were recruited in the study. Out of those 237 subjects, 163 (100, 61.3% for the BPH group and 63, 38.7% for the control group) completed the baseline interview. Compared to the control group, subjects with BPH were older (mean age of 69.1, SD 9.09 vs. mean age of 57.1, SD 12.6, p -value <0.01), were less likely to be actively employed (26% vs. 54.0%, p -value <0.01); had a lower household income (80.4% vs. 54.8%, p -value <0.01), were less likely to smoke (85.0% vs. 65.1%, p -value <0.01), and had had a stroke (6.0% vs. 0%, p -value <0.05). The baseline characteristics of study subjects are shown in **Table 1**.

Descriptive statistics and floor and ceiling effects

Table 2 shows the descriptive statistics of SF-12v2, IPSS and DASS-21 scores. With regards to the IPSS total symptom score and IPSS-S, 1.89% of the subjects had the lowest possible score whilst 0.63% of subjects had the highest possible score. With regards to the IPSS-V, it had a relatively higher floor effect (21.6%) but a lower ceiling effect (0.62%). With regards to the IPSS HRQOL score, 1.84% of subjects had the lowest possible score while 4.29% of subjects had the highest possible score.

Construct validity

Table 3 shows the results of the corrected item-total correlation testing for the items regarding urinary symptoms. Corrected item-total correlation scores were ≥ 0.4 for all symptoms, except in the case of nocturia (correlation coefficient: 0.309). Pearson's correlation testing of the IPSS total symptom score against clinical parameters (voided volume, post-void residual urine and Qmax), and the IPSS HRQOL score against the SF-12v2 and DASS-21 scores were then tested. The correlations between the IPSS total symptom score and clinical parameters were not statistically significant. On the other hand, there were significant correlations between the IPSS HRQOL score, SF-12v2 and DASS-21 scores. However, the IPSS HRQOL score was weakly correlated with SF-12v2 PCS (correlation coefficient: -0.30) and MCS (correlation coefficient: -

0.26), and the DASS-21 Depression (correlation coefficient: 0.32), Anxiety (correlation coefficient: 0.31) and Stress (correlation coefficient: 0.30) scores respectively.

Reliability

The internal consistency and test-retest reliability of the IPSS were then performed. With regards to the IPSS total symptom score, it increased from 13.61 to 14.24 while the HRQOL score decreased from 3.11 to 2.99 from the baseline. Both are considered insignificant. Cronbach's alpha coefficient was 0.90 for the seven symptom-related items. The ICC of the IPSS total symptom score and HRQOL score were 0.90 and 0.86, respectively.

Sensitivity

A statistically significant difference was detected between the BPH group and non-BPH group for the IPSS total symptom score (effect size 0.68) but not the HRQOL score. Moreover, there was no significant difference between groups for all the SF-12v2 and DASS-21 scores.

Furthermore, the area under the ROC curve was 0.67 (95% CI: 0.584, 0.757) and 0.60 (95% CI: 0.506, 0.689) with regards to the IPSS total symptom score and HRQOL score, respectively.

Figure 1 depicts the area under the ROC curve and the discriminative power of the IPSS total symptom and HRQOL scores in the differentiation of the BPH and control groups. **Figure 2** shows the ability of the IPSS scores, SF-12v2 scores and DASS-21 scores when discriminating between benign prostatic hyperplasia and control patients by the independent t-test.

Responsiveness

Table 4 shows the results of the responsiveness analysis using the paired-t-test and responsiveness statistics. There were no significant differences between the IPSS total symptom score, IPSS-V, IPSS-S, IPSS-V/S and HRQOL scores; SF-12v2 PCS and MCS scores; and DASS-21 Depression, Anxiety and Stress scores.

Table 5 shows the responsiveness properties of the HRQOL instruments in a subgroup of patients with improved health status over a 6-month period. Improvements in the IPSS total symptom score, and the IPSS-V, IPSS-S and HRQOL scores were statistically significant and

corresponded to a medium-sized change effect. Changes in the IPSS-V/S; SF-12v2 PCS and MCS scores, and the DASS-21 Depression, Anxiety and Stress scores were small-sized and therefore statistically insignificant.

Discussion

The purpose of this study is to evaluate psychometric properties including validity, reliability, sensitivity and responsiveness of the IPSS instrument in male patients with BPH under secondary care. Although this is not the first study to validate the IPSS in a Chinese population, the current study was designed to build upon previous research to provide more information on the psychometric properties of the IPSS.

The lack of significant floor and ceiling effects of the IPSS indicated that this measure has the potential capability to capture any deterioration or improvement of LUTS and HRQOL.

In item-total correlation testing, the symptom-related item of nocturia had a poor correlation. The result suggest that this item is being measured in a related yet slightly different domain than the other items of the IPSS. This result is similar to that of a previous validation study carried out on subjects under primary care in Hong Kong [15].

Contrary to the hypothesis, the IPSS total symptom score did not present any statistically significant correlations with objective parameters including Qmax, voided volume and post-void residual urine. The findings implied that there were distinct differences between objective clinical parameters and perceived symptom severity. Therefore, both clinical parameters and the IPSS should be combined and supplement one another. These findings have supplemented the previous Hong Kong study [15] which did not test the correlation between the IPSS total symptom score and objective clinical parameters.

The IPSS HRQOL score had a weak correlation with SF-12v2 PCS and MCS scores suggesting that there is a difference between the construct of the IPSS HRQOL question and that of the SF-12v2. Since SF-12v2 is a generic HRQOL measure, the domains of such measures might not be specific nor sensitive enough to capture the impact of LUTS on HRQOL. Generic measures

contain irrelevant domains and may miss specific concerns of the respondents [32]. The results support the added value of condition-specific measures.

The IPSS was found to be a reliable measure for patient-reported outcome. The internal consistency (Cronbach's alpha >0.7) of the IPSS appears to be comparable to the original English instrument [8]. The 2-week test-retest reliability was acceptable and comparable to other versions of the IPSS [8, 15, 26, 33].

The findings of higher severity IPSS total symptoms severity in patients with BPH compared to the non-BPH group showed that the total symptom score was sensitive in differentiating patients. The area under the ROC curve was just short of 0.7 (0.67) but still acceptable. Unexpectedly, the IPSS HRQOL question was not sensitive enough to differentiate subjects. The single item question seemingly does not suffice. Thus, clinicians and researchers should not rely on the IPSS single-item HRQOL for screening. Expectedly, the SF-12v2 and DASS-21 was not sensitive enough to differentiate patients due to the generic nature of these measures.

Limitations

Subjects in the present study were only recruited by a convenience sampling in three government funded urology specialist clinics. The psychometric performance of these measures should be further tested with Putonghua speakers or Chinese speakers of different nationalities as the subjects in this study were mainly Cantonese speakers.

Conclusion

The IPSS is a valid and reliable measure among Chinese patients with BPH. The IPSS total symptom score, but not the HRQOL score, is sensitive in differentiating subgroups.

Acknowledgements

The researchers would like to express their deepest gratitude to Dr James Tsu and Mr Charles Wong, Department of Surgery of QMH, Dr Peggy Chu and Dr Alan Ip, Department of Surgery of TMH and Dr Fu-keung Cheung, Department of Surgery of PMH, for the coordinating data collection and subject recruitment. Much appreciation goes to the research assistants who assisted in this project including Eileen Yeung (data collection) and Ryan Pak (data analysis).

Running Title: IPSS in Chinese male with BPH

Financial disclosure

This study was funded by the Hong Kong Urological Association Council. The funders had no role in the design, data collection or analysis of this study.

Figure Legend

Figure 1. Receiver operating characteristics curves of the IPSS total symptom and HRQOL scores

Figure 2. Ability of IPSS scores, SF-12v2 scores and DASS-21 scores in discriminating between benign prostatic hyperplasia and control patients by the independent t-test

References

1. Sarma, A.V. and J.T. Wei, *Clinical practice. Benign prostatic hyperplasia and lower urinary tract symptoms*. N Engl J Med, 2012. **367**(3): p. 248-57.
2. Wong, S.Y., et al., *Risk factors for lower urinary tract symptoms in southern Chinese men*. Urology, 2006. **68**(5): p. 1009-14.
3. Irwin, D.E., et al., *Population-based survey of urinary incontinence, overactive bladder, and other lower urinary tract symptoms in five countries: results of the EPIC study*. European urology, 2006. **50**(6): p. 1306-1315.
4. Welch, G., K. Weinger, and M.J. Barry, *Quality-of-life impact of lower urinary tract symptom severity: results from the Health Professionals Follow-up Study*. Urology, 2002. **59**(2): p. 245-50.
5. Trueman, P., et al., *Prevalence of lower urinary tract symptoms and self-reported diagnosed 'benign prostatic hyperplasia', and their effect on quality of life in a community-based survey of men in the UK*. BJU Int, 1999. **83**(4): p. 410-5.
6. Tsang, K.K. and W.M. Garraway, *Impact of benign prostatic hyperplasia on general well-being of men*. Prostate, 1993. **23**(1): p. 1-7.
7. Ozturk, M.I., et al., *International prostate symptom score: really appreciated by all patients or not?* Urol J, 2011. **8**(3): p. 227-30.
8. Barry, M.J., 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): p. 1549-57; discussion 1564.
9. van Venrooij, G.E., T.A. Boon, and R.P. de Gier, *International prostate symptom score and quality of life assessment versus urodynamic parameters in men with benign prostatic hyperplasia symptoms*. J Urol, 1995. **153**(5): p. 1516-9.
10. Asiedu, B., et al., *The role of sex steroid hormones in benign prostatic hyperplasia*. Aging Male, 2017. **20**(1): p. 17-22.
11. Yassin, A., et al., *Effects of testosterone replacement therapy withdrawal and re-treatment in hypogonadal elderly men upon obesity, voiding function and prostate safety parameters*. Aging Male, 2016. **19**(1): p. 64-9.

12. Kim, M.K., et al., *Relationship of sex hormones and nocturia in lower urinary tract symptoms induced by benign prostatic hyperplasia*. *The Aging Male*, 2012. **15**(2): p. 90-95.
13. Park, S.G., et al., *Impact of metabolic status on the association of serum vitamin D with hypogonadism and lower urinary tract symptoms/benign prostatic hyperplasia*. *Aging Male*, 2017: p. 1-5.
14. Aktas, B.K., et al., *Impact of metabolic syndrome on erectile dysfunction and lower urinary tract symptoms in benign prostatic hyperplasia patients*. *Aging Male*, 2011. **14**(1): p. 48-52.
15. Choi, E.P., C.L. Lam, and W.-Y. Chin, *Validation of the International Prostate Symptom Score in Chinese males and females with lower urinary tract symptoms*. *Health and quality of life outcomes*, 2014. **12**(1): p. 1.
16. Choi, E.P., et al., *The responsiveness of the International Prostate Symptom Score, Incontinence Impact Questionnaire-7 and Depression, Anxiety and Stress Scale-21 in patients with lower urinary tract symptoms*. *Journal of advanced nursing*, 2015. **71**(8): p. 1857-1870.
17. Chin, W.Y., E.P. Choi, and C.L. Lam, *The effect of timing of incentive payments on response rates for cohort study telephone interviews in primary care setting with cost-minimization analysis, a randomized controlled trial*. *BMC medical research methodology*, 2015. **15**(1): p. 79.
18. Ware, J.E., M. Kosinski, and S.D. Keller, *SF-12: How to score the SF-12 physical and mental health summary scales*. 1995: Health Institute, New England Medical Center.
19. Ware, J.E., et al., *How to score version 2 of the SF-12 health survey (with a supplement documenting version 1)*. 2002: QualityMetric Incorporated.
20. Lam, E.T., et al., *Is the SF-12 version 2 Health Survey a valid and equivalent substitute for the SF-36 version 2 Health Survey for the Chinese?* *J Eval Clin Pract*, 2013. **19**(1): p. 200-8.
21. Lam, C.L., et al., *Population norm of Chinese (HK) SF-12 health survey-version 2 of Chinese adults in Hong Kong*. *Hong Kong Practitioner*, 2010. **32**(2): p. 77-86.
22. Gloster, A.T., et al., *Psychometric properties of the Depression Anxiety and Stress Scale-21 in older primary care patients*. *J Affect Disord*, 2008. **110**(3): p. 248-59.

23. Henry, J.D. and J.R. Crawford, *The short-form version of the Depression Anxiety Stress Scales (DASS-21): construct validity and normative data in a large non-clinical sample*. Br J Clin Psychol, 2005. **44**(Pt 2): p. 227-39.
24. Lam, T.P., et al., *Psychological well-being of interns in Hong Kong: what causes them stress and what helps them*. Med Teach, 2010. **32**(3): p. e120-6.
25. Ownsworth, T., et al., *Assessing emotional status following acquired brain injury: the clinical potential of the depression, anxiety and stress scales*. Brain Inj, 2008. **22**(11): p. 858-69.
26. Badia, X., et al., *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): p. 614-20.
27. Terwee, C.B., et al., *Quality criteria were proposed for measurement properties of health status questionnaires*. J Clin Epidemiol, 2007. **60**(1): p. 34-42.
28. Ware, J.E., Jr. and B. Gandek, *Methods for testing data quality, scaling assumptions, and reliability: the IQOLA Project approach*. International Quality of Life Assessment. J Clin Epidemiol, 1998. **51**(11): p. 945-52.
29. Nunnally, J.C., *Psychometric theory*. 1978: McGraw-Hill.
30. Guyatt, G., S. Walter, and G. Norman, *Measuring change over time: Assessing the usefulness of evaluative instruments*. Journal of Chronic Diseases, 1987. **40**(2): p. 171-178.
31. Liang, M.H., A.H. Fossel, and M.G. Larson, *Comparisons of five health status instruments for orthopedic evaluation*. Medical care, 1990. **28**(7): p. 632-642.
32. Patrick, D.L. and R.A. Deyo, *Generic and disease-specific measures in assessing health status and quality of life*. Medical care, 1989: p. S217-S232.
33. Quek, K.F., et al., *Reliability and validity of the Malay version of the International Prostate Symptom Score in the Malaysian population*. J Urol, 2002. **167**(3): p. 1359-62.

Table 1. Baseline characteristics at baseline between BPH participants and control patients

	Total (N = 163)	BPH patients (N = 100)	Control Patients (N = 63)	P-value
Socio-Demographic (% , n)				
Age (mean ± SD)	64.50±12.05 (163)	69.14±9.09 (100)	57.14±12.55 (63)	<0.001*
Marital Status				0.383
Not Married	8.70 % (14)	7.14 % (7)	11.11 % (7)	
Married	91.30 % (147)	92.86 % (91)	88.89 % (56)	
Employment Status				<0.001*
Not Working	63.19 % (103)	74.00 % (74)	46.03 % (29)	
Currently working	36.81 % (60)	26.00 % (26)	53.97 % (34)	
Household Income				<0.001*
< \$20,000	70.13 % (108)	80.43 % (74)	54.84 % (34)	
≥ \$20,000	29.87 % (46)	19.57 % (18)	45.16 % (28)	
Smoking Status				0.003*
Non-Smoker	77.30 % (126)	85.00 % (85)	65.08 % (41)	
Smoker	22.70 % (37)	15.00 % (15)	34.92 % (22)	
Drinking Status				0.736
Non-Drinker	64.97 % (102)	65.98 % (64)	63.33 % (38)	
Drinker	35.03 % (55)	34.02 % (33)	36.67 % (22)	
Voided volume (mean ± SD)	245.31±128.73 (81)	245.31±128.73 (81)	NA	NA
Post void residual urine (mean ± SD)	86.36±91.08 (81)	86.36±91.08 (81)	NA	NA
Qmax (mean ± SD)	12.61±7.19 (77)	12.61±7.19 (77)	NA	NA
Co-morbidities (% , n)				
Hypertension	52.47 % (85)	49.00 % (49)	58.06 % (36)	0.261
Diabetes	16.15 % (26)	16.16 % (16)	16.13 % (10)	0.996
Cardiovascular Disease	20.00 % (32)	24.49 % (24)	12.90 % (8)	0.074
Stroke	3.68 % (6)	6.00 % (6)	0.00 % (0)	0.048*
Respiratory diseases	7.45 % (12)	6.12 % (6)	9.52 % (6)	0.423
Cancer	4.91 % (8)	5.00 % (5)	4.76 % (3)	0.945
Kidney diseases	5.88 % (9)	7.00 % (7)	3.77 % (2)	0.420
Psychiatric disorders	4.35 % (7)	3.03 % (3)	6.45 % (4)	0.300
Other major diseases	25.15 % (41)	22.00 % (22)	30.16 % (19)	0.242

BPH= benign prostatic hyperplasia; Qmax=maximum urinary flow rate; NA=Not applicable

Note:

* Significant at a level of 0.05 by chi-square test or independent t-test as appropriate

Table 2. Descriptive statistics of IPSS, SF-12v2 scores and DASS-21 scores at baseline

	Total	Floor (%)	Ceiling (%)
IPSS scores (mean ± SD, n)			
IPSS Total Symptom Score (min 0 - max 35)	13.03 ± 8.09 (159)	1.89%	0.63%
IPSS Voiding subscore (min 0 - max 20)	6.58±5.54 (162)	21.60%	0.62%
IPSS Storage subscore (min 0 - max 15)	6.34±3.58 (159)	1.89%	0.63%
IPSS Voiding-to-storage subscore ratio	1.17±1.15 (156)		NA
IPSS HRQOL Score (min 0 - max 6)	2.99 ± 1.69 (163)	1.84%	4.29%
SF-12v2 Scores (mean ± SD, n)			
PCS	44.05 ± 11.20 (163)	0.61%	0.61%
MCS	53.00 ± 11.18 (163)	0.61%	0.61%
DASS-21 scores (mean ± SD, n)			
DASS Depression Score (min 0 - max 21)	2.01 ± 3.22 (163)	47.85%	0.61%
DASS Anxiety Score (min 0 - max 21)	2.28 ± 2.53 (163)	25.77%	0.61%
DASS Stress Score (min 0 - max 21)	2.44 ± 3.34 (163)	41.72%	0.61%

PCS = Physical Component Summary; MCS = Mental Component Summary; MH = Mental Health;

HRQOL = health-related quality of Life; IPSS = International Prostate Symptom Score;

DASS = Depression Anxiety Stress Scales; NA = Not Applicable

Table 3. Corrected item-scale correlation of the IPSS

	Eligible Patients (N=163)		
	n	mean \pm SD	Corrected item-total correlation
IPSS symptom score			
Incomplete emptying	162	2.062 \pm 1.806	0.577
Frequency	162	2.438 \pm 1.642	0.517
Intermittency	162	1.556 \pm 1.759	0.589
Urgency	162	1.667 \pm 1.712	0.583
Weak Stream	163	1.681 \pm 1.797	0.651
Straining	163	1.294 \pm 1.732	0.553
Nocturia	161	2.199 \pm 1.382	0.309

IPSS = International Prostate Symptom Score

Table 4. Responsiveness to change over time from baseline to 6-month follow-up (N=91)

	Baseline	6-month follow-up	Mean Change \pm SD	P-value	SES	SRM
IPSS (mean \pm SD)						
Total Symptom Score	13.66 \pm 7.62	12.67 \pm 8.28	-0.99 \pm 6.60	0.161	0.13	0.15
Voiding subscore	6.94 \pm 5.38	6.62 \pm 5.80	-0.32 \pm 4.41	0.490	0.06	0.07
Storage subscore	6.64 \pm 3.38	5.98 \pm 3.64	-0.66 \pm 3.17	0.052	0.20	0.21
Voiding-to-storage subscore ratio	1.11 \pm 0.90	1.25 \pm 1.19	0.15 \pm 0.96	0.158	0.16	0.15
HRQOL Score	3.01 \pm 1.72	2.77 \pm 1.75	-0.24 \pm 1.64	0.162	0.14	0.15
SF-12v2 (mean \pm SD)						
PCS	43.64 \pm 10.90	43.16 \pm 12.62	-0.48 \pm 8.42	0.585	0.04	0.06
MCS	52.41 \pm 11.86	52.02 \pm 12.19	-0.39 \pm 11.18	0.743	0.03	0.03
DASS-21 (mean \pm SD)						
DASS Depression Score	2.27 \pm 3.42	2.43 \pm 3.70	0.15 \pm 2.00	0.465	0.04	0.08
DASS Anxiety Score	2.54 \pm 2.57	2.46 \pm 2.71	-0.08 \pm 1.79	0.683	0.03	0.04
DASS Stress Score	2.66 \pm 3.69	2.65 \pm 3.69	-0.01 \pm 2.07	0.960	0.00	0.01

IPSS = International Prostate Symptom Score; DASS = Depression Anxiety Stress Scales;

SES = Standardized Effect Size; SRM = Standardized Response Mean;

PCS = Physical Component Summary; MCS = Mental Component Summary

Note:

*Significant difference (P<0.05) on score between baseline and 6-month follow-up by paired t-test

Table 5. Responsiveness to change over time from baseline to 6-month follow-up among patients with improved health status (N=27)

	Baseline	6-month follow-up	Mean Change \pm SD	P-value	SES	SRM
IPSS (mean \pm SD)						
Total Symptom Score	15.35 \pm 8.42	10.81 \pm 8.11	-4.54 \pm 8.28	0.010*	0.54	0.55
Voiding subscore	7.88 \pm 5.29	5.12 \pm 5.35	-2.77 \pm 5.19	0.012*	0.52	0.53
Storage subscore	7.46 \pm 3.57	5.69 \pm 3.46	-1.77 \pm 3.88	0.029*	0.50	0.46
Voiding-to-storage subscore ratio	0.99 \pm 0.53	0.84 \pm 0.85	-0.15 \pm 0.97	0.432	0.29	0.16
HRQOL Score	3.22 \pm 1.85	2.48 \pm 1.63	-0.74 \pm 1.81	0.043*	0.40	0.41
SF-12v2 (mean \pm SD)						
PCS	42.91 \pm 11.68	44.51 \pm 11.71	1.6 \pm 9.53	0.391	0.14	0.17
MCS	50.96 \pm 12.47	50.78 \pm 13.26	-0.17 \pm 12.68	0.944	0.01	0.01
DASS-21 (mean \pm SD)						
DASS Depression Score	2.56 \pm 3.53	2.48 \pm 4.16	-0.07 \pm 2.30	0.869	0.02	0.03
DASS Anxiety Score	3.04 \pm 2.98	2.78 \pm 3.25	-0.26 \pm 2.40	0.579	0.09	0.11
DASS Stress Score	3.04 \pm 3.75	2.96 \pm 4.32	-0.07 \pm 2.56	0.881	0.02	0.03

IPSS = International Prostate Symptom Score; DASS = Depression Anxiety Stress Scales;

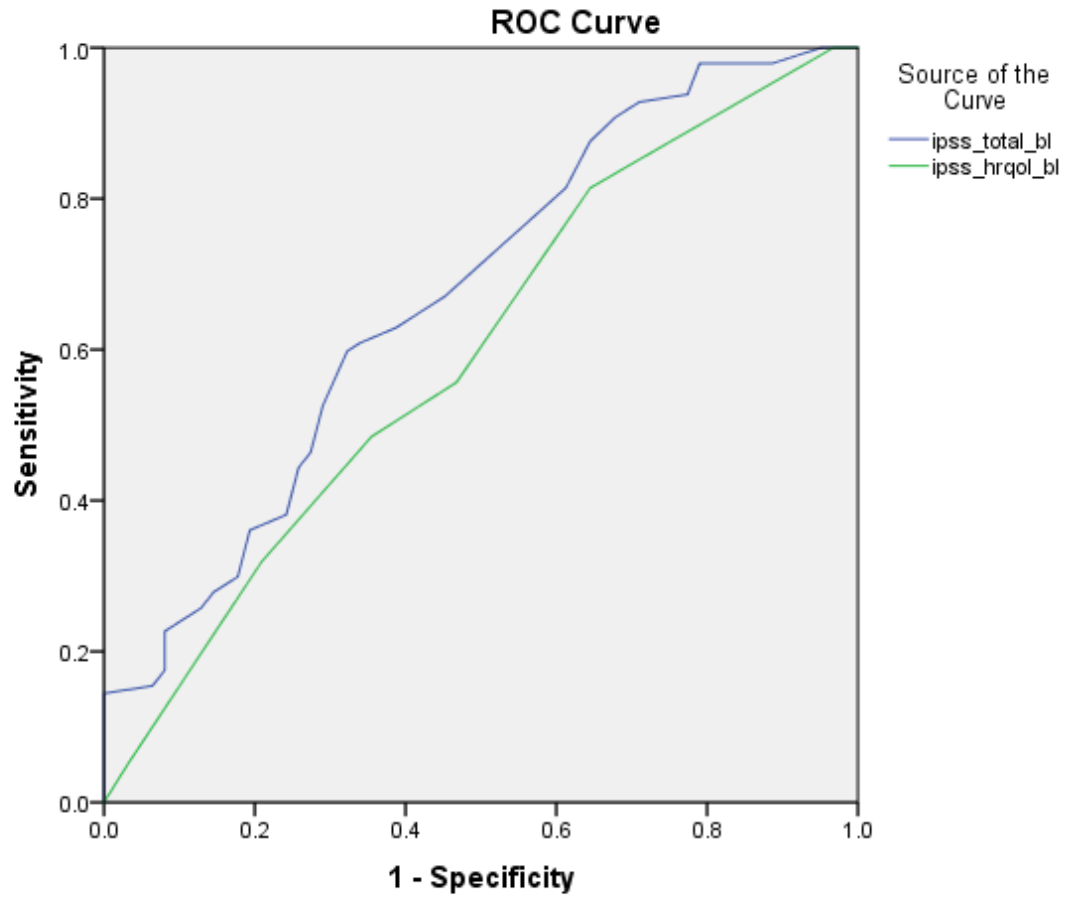
SES = Standardized Effect Size; SRM = Standardized Response Mean;

PCS = Physical Component Summary; MCS = Mental Component Summary

Note:

*Significant difference (P<0.05) on score between baseline and 6-month follow-up by paired t-test

Figure 1. ROC curves of IPSS total symptom and HRQOL scores



Diagonal segments are produced by ties.

Figure 2. Ability of IPSS scores, SF-12v2 scores and DASS-21 scores in discriminating between benign prostatic hyperplasia and control patients by the independent t-test

