Title	The internal and external responsiveness of Functional Assessment of Cancer Therapy-Prostate (FACT-P) and Short Form-12 Health Survey version 2 (SF-12 v2) in patients with prostate cancer
Author(s)	Choi, EPH; Wong, CKH; Wan, EYF; Tsu, HLJ; Chin, WY; Kung, K; Yiu, MK
Citation	Quality of Life Research, 2016, v. 25 n. 9, p. 2379-2393
Issued Date	2016
URL	http://hdl.handle.net/10722/223856
Rights	The final publication is available at Springer via http://dx.doi.org/10.1007/s11136-016-1254-1; This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

Title page

The internal and external responsiveness of Functional Assessment of Cancer Therapy-Prostate (FACT-P) and Short Form-12 Health Survey version 2 (SF-12 v2) in Patients with Prostate Cancer

Order of Author: Edmond P.H. Choi[1*], Carlos K.H. Wong[2], Eric Y.F Wan[2], James H.L. Tsu[3], W.Y. Chin[2], Kenny Kung[2], M.K. Yiu[3]

[1] School of Nursing, University of Hong Kong, Pok Fu Lam, Hong Kong

[2] Department of Family Medicine and Primary Care, University of Hong Kong, Ap Lei Chau, Hong Kong

[3] Division of Urology, Department of Surgery, University of Hong Kong, Pokfulam, Hong Kong

* Corresponding author

Corresponding Author Information:

Edmond P.H. Choi[1]

School of Nursing

Address: 4/F, William M.W. Mong Block, 21 Sassoon Road, Pokfulam, Hong Kong

Tel: (+852) 3917 6677

Email: h0714919@connect.hku.hk

Acknowledgements

The authors wish to express their gratitude to Professor Cindy L.K Lam for design and planning of this study and Charles Wong for his assistance in data collection.

Running Title: Responsiveness of HRQOL measures for Prostate Cancer

Keywords: Prostate cancer; Responsiveness; Health-related quality of life, instrument; SF-12; FACT-P; Chinese

Abstract

Purpose: To examine the responsiveness of Functional Assessment of Cancer Therapy-Prostate (FACT-P) and Short Form-12 Health Survey version 2 (SF-12 v2) in prostate cancer patients because there is a lack of evidence to support their responsiveness in prostate cancer patients.

Methods: One hundred sixty-eight subjects with prostate cancer were surveyed at baseline and at 6 months using the SF-12 v2 and FACT-P version 4. Internal responsiveness was assessed using paired t-test and generalized estimating equation. External responsiveness was evaluated using receiver operating characteristic curve analysis.

Results: The internal responsiveness of the FACT-P and SF-12 v2 to detect positive change was satisfactory. The FACT-P and SF-12 v2 could not detect negative change. The FACT-P and the SF-12 v2 performed the best in distinguishing between improved general health and worsened general health. The FACT-P performed better in distinguishing between unchanged general health and worsened general health. The SF-12 v2 performed better in distinguishing between unchanged general health and improved general health.

Conclusions: Positive change detected by these measures should be interpreted with caution as they might be too responsive to detect "noise", which is not clinically significant. The ability of the FACT-P and the SF-12 v2 to detect negative change was disappointing. The internal and external responsiveness of the Social Well-Being of the FACT-P cannot be supported, suggesting that it is not suitable to longitudinally monitor the social component of HRQOL in prostate cancer patients. The study suggested that generic and disease–specific measures should be used together to complement each other.

Words Count: 250 out of 250

Manuscript Text

1. Introduction

According to the worldwide burden of cancer study in 2008, prostate cancer is the second most common cancer (in terms of incidence) in adult males worldwide [1]. The reported incidence rates of prostate cancer were higher in western countries than other countries most likely because of the widespread use of prostate specific antigen (PSA) testing and biopsy in these developed countries [1]. The incidence rates of prostate cancer has been rising in most Asian countries, thought to be due to westernization of the lifestyle [2]. Through early diagnosis by PSA testing and advanced treatment modalities, patients with localized prostate cancer nowadays can achieve survival rates close to 100 % [3]. With increased detection of prostate cancer and associated advancements in medical technology, changes in the epidemiology of the disease are being observed with more diagnoses occurring in younger males and at earlier stages of the disease. Prostate cancer patients now live longer with their disease. As the health-related quality of life (HRQOL) of patients with prostate cancer are often affected either as a direct result of the cancer itself or from side effects of interventions [4-8], HRQOL is one of the important clinical outcomes in patients with prostate cancer [9]. Many clinical trials include HRQOL as one of the outcome measures in order to fully evaluate the effectiveness of prostate cancer treatments [9] and there is significant interest amongst oncologists and epidemiologists to explore the trajectories of HRQOL in cancer patient populations in order to understand the factors which may influence the change in HRQOL over time.

Many different HRQOL measures are available for prostate cancer patients [9-11]. A recent systematic review evaluated the twenty most commonly used HRQOL measures for studies on prostate cancer [10]. Among different generic HRQOL measures, the review only recommended the Short Form-12 (SF-12) because of its well established psychometric properties [10]. Besides, the SF-12 is widely used in general populations and different patient populations, worldwide. The review by Haemoen also recommended the Functional Assessment of Cancer Therapy-Prostate (FACT-P) for the specific evaluation of prostate cancer patients' HRQOL because the measure has good psychometric properties [10]. The FACT-P consists of prostate cancer subscale and the core module of the Functional Assessment of Chronic Illness Therapy (FACIT), which is applicable to all cancer patient populations.

HRQOL measures should have well-established psychometric properties before application in clinical settings and research studies [12]. Responsiveness is one of the most important psychometric properties for studies monitoring cohorts over time because responsiveness will impact the interpretation of findings and subsequently the conclusions in longitudinal studies [13]. The use of a HRQOL measure that is not responsive can lead to type II error (false negative) in clinical trials [14] and inaccurate estimation of HRQOL trajectories in longitudinal observational studies. Therefore, the responsiveness of HRQOL measures should be confirmed before they are applied in longitudinal studies. However, there is still a lack of evidence on the responsiveness of the SF-12 v2 and FACT-P in patients with prostate cancer.

Two distinct methods can be used to assess the responsiveness of an instrument: internal responsiveness (distribution-based) and external responsiveness (anchor-based). Internal responsiveness is the ability of an instrument to detect a clinically important change over time induced by an intervention that has been shown to be effective. According to Husted et al, paired t-test with effect size statistics was commonly used to evaluate the internal responsiveness. External responsiveness refers to the ability of an instrument to detect a clinically important change over time with reference to an external standard for health status (external anchor) shown by receiver operating characteristic (ROC) curves [13, 15-17].

The aims of this study were to examine the internal and external responsiveness of the FACT-P and SF-12 v2 measures in Chinese patients with prostate cancer.

2. Methods

2.1 Subject and Study Design

This was a prospective longitudinal study. Convenience sampling of Chinese patients with confirmed diagnosis of prostate cancer were recruited in a urological specialist outpatient clinic of a teaching hospital in Hong Kong between May 2013 and January 2014. Patients were excluded if they (i) refused to give consent, (ii) had hearing problems, (iii) were not able to communicate in Chinese/Cantonese, or (vi) being too ill to complete the questionnaire. Subjects who consented were asked to provide their contact details and were subsequently interviewed by a trained interviewer who administered the study questionnaire by a face-to-face interview (baseline). Subjects were contacted again at 6-month after their baseline interview to complete a follow-up telephone interview. The interviewer was required to read the study questionnaire verbatim in a standardized interview approach. We deliberately used

interviewer-administered method because many of our subjects were the elderly with poor literacy level. They were not able to complete the questionnaire themselves. Interviewer administration can enhance response rates and reduce missing values [18].

2.2 Study instruments

The Chinese (Hong Kong) Short Form-12 version 2 (SF-12v2)

The SF-12v2 Health Survey is a generic HRQOL measure with eight subscales (Physical functioning, PF; Role physical, RP; Bodily pain, BP; General health, GH; Vitality, VT; Social functioning, SF; Role emotional, RE; Mental health, MH) and two summary scales (Physical composite summary, PCS-12; Mental composite summary, MCS-12). The possible range of each domain score is 0-100. The two summary scores are norm-based scoring with the population mean of 50 and standard deviation of 10. A high score indicates better HRQOL [19, 20]. This generic instrument is shown to be valid and reliable in Chinese population [21].

The Functional Assessment of Cancer Therapy-Prostate version 4 (FACT-P version 4)

FACT-P (version 4) is an extension of the FACT-G HRQOL instrument that emphasizes on a range of important aspects of HRQOL specific to patients with prostate cancer. It has 39 items that are categorized into four FACT-G subscales (Physical well-being, PWB; Social well-being, SWB; Emotional well-being, EWB; Functional well-being FWB) and one Prostate Cancer Subscale (PCS) addressing the additional concerns about prostate cancer [22]. The raw scores are computed to give standard scores in the possible range of 0-28 for the PWB, EWB and FWB subscales, 0-24 for the SWB subscale, and 0-48 for the PCS. A Trial Outcome Index (TOI) is the total of PWB, FWB and PCS whereas FACT-P total score is the sum of all five subscales with a range of 0-104 for TOI score, 0-108 for FACT-G Total score, and 0-156 for FACT-P Total score. Higher scores in subscales and total score indicate better HRQOL. The FACT-P (version 4) is a valid, reliable and sensitive measure to assess the HRQOL of Chinese patients with prostate cancer [23].

Global Rating of Change Scale (GRS)

The GRS was used as an external anchor because a review paper reported that the GRS was typically used in studies which evaluated the responsiveness of instruments [24]. Furthermore, the GRS had good face validity, construct validity, test-retest reliability [24]. A previous

study which evaluated the internal and external responsiveness of Functional Assessment of Cancer Therapy-Colorectal and the SF-12 v2 also used the GRS as an external anchor [25]. All subjects were asked this single item scale during their 6-month follow-up interview to evaluate their subjective changes in global health condition by a retrospective question "Compared to the first visit (six months ago), how would you rate your overall health now?" [26]. The response was rated on a 7-point ordinal scale ranging from -3 to 3 anchored from the much worse to the much better options, with 0 indicating no change. The GRS has been commonly used as the external criterion to estimate minimal clinically important difference and responsiveness of measures [25, 27-33].

2.3 Statistical Analysis

Responsiveness was assessed using the self-reported health change anchor to define samples reflecting "unchanged" (rating of 0), "improved" (rating of 1 to 3) and "worsened" (rating of -3 to -1) in health status [15]. Mean and standard deviation of all HRQOL scores were calculated for each sample.

Internal responsiveness

For assessing the internal responsiveness, mean changes over the past six months were tested by paired t-test in patients who were classified as "worsened", "unchanged" and "Improved". The HRQOL score differences between baseline and follow-up assessments were also tested using the standardized effect size (SES) [26], standardized response mean (SRM) [34] and responsiveness statistics (RS) [29, 35, 36] separately for each groups. Three responsiveness statistics were reported because the method for calculating the most appropriate responsiveness statistic was still controversial [17]. Appendix 1 shows the equations of these three responsiveness statistics

The value of SES, SRM and RS were interpreted as trivial for <0.2, small for ≥ 0.2 and <0.5, moderate for ≥ 0.5 and <0.8 or large for ≥ 0.8, according to criteria defined by Cohen [37], Liang [34] and Norman [38] respectively. Internal responsiveness was supported if these changes were interpreted as small or above. Thus, 95% bootstrap bias-corrected and accelerated confidence intervals [39] for SES, SRM and RS were obtained using the bootstrapping estimation method with 2000 replications.

Sensitivity analysis was conducted using generalized estimating equation (GEE) model by Global Rating on Change Scale after multiple imputation, controlling for baseline socio-

demographic and clinical characteristics. Multiple imputation was used to handle the missing date. In this sensitivity analysis, each missing value was imputed by the chained equation method 20 times. For each of the 20 imputed datasets, the same analysis was performed with the twenty sets of results combined using Rubin's combination rules [40]. To confirm the internal responsiveness of the instruments over times accounting for within-subject correlation with repeated measurements, GEE assigning time as dependent variables with an identity link function was conducted.

External responsiveness

Based on the recommendation by Husted et al [17] and Deyo et al[41], we evaluated the external responsiveness of these measures by receiver operating characteristic (ROC) curve analysis. However, it should be noted that one of the major disadvantages of the ROC method is that the external anchor must be dichotomized. Therefore, we had three different comparisons: (i) improved vs. unchanged group, (ii) improved vs. worsened group, and (iii) worsened vs. unchanged group. First, independent t-tests were performed to compare the mean changes between groups. Subsequently, the ROC curve analysis was performed to assess the ability of these measures to detect HRQOL score change with health condition changes or to discriminate between groups. The ROC curve is a plot of the true-positive rate (sensitivity) against the false-positive rate (1-specificity). Conceptually, the ROC curve can provide an overview of the relationship between a measure and an external anchor. The area under a ROC curve (AUC) can show the probability that a measure correctly classified patients according to the external anchor. Perfect discriminatory power is defined as a value of 1 but a value of 0.5 is considered no discriminatory power. To support the external responsiveness, the AUC was considered adequate for ≥ 0.7 [17, 42] and its 95% confidence intervals were reported.

All statistical analyses were conducted by the Stata 13 (StataCorp, College Station, TX) with p-values < 0.05 indicating statistical significance.

3. Results

3.1 Baseline Characteristics

A total of 339 patients with prostate cancer were invited to join the study. Of the 339 patients, 29 patients refused to participate. A further 19 patients were excluded because of hearing problems, inability to communicate in Chinese/Cantonese, or being too ill to complete the

questionnaire. Among them, 291 eligible subjects (response rate: 85.9 %) completed baseline interviews. Of the 291 subjects, 168 subjects completed 6-month follow-up interviews (attrition rate: 42.3%). Subject recruitment flowchart is shown in figure 1. Baseline characteristics of defaulted and followed subjects are shown in table 1. The age of our subjects at baseline ranged from 41 to 99 years old. There were statistically significant differences in mean age, educational attainment, marital status, Karnofsky performance status, treatment (radical prostatectomy), all FACT-P scores (except EWB) and all SF-12 v2 scores (except MCS-12) between subjects who completed 6-month telephone follow-up interviews and those who did not.

Baseline and 6-month follow-up on the HRQOL scores are shown in Table 2. In both baseline and 6-month follow-up interviews, the means of the FACT-G and FACT-P total score were greater than 80% of the maximum possible score for each scale, respectively. The mean change of the PWB, EWB, TOI and Prostate Cancer Subscale, FACT-G total scale and FACT-P was statistically significant between the two time points (P-value < 0.001). In regard to the SF-12 v2, the baseline PCS-12 score was lower than 50 (the population mean) whilst the bassline MCS-12 score was higher than 50. The mean change of the RP, BP, VT, RE and MH and MCS-12 was statistically significant between the two time points (P-value< 0.001).

Table 3 shows that most subjects reported no change (71.43%) in the GRS. 8.34% and 20.24% of subjects rate “deterioration” and “improvement” in current general health conditions, compared to six months ago, respectively.

3.2 Internal Responsiveness

A summary of the mean change and responsiveness statistics for each group (i.e. worsened, unchanged and improved) is shown in Table 4. The results of sensitivity analysis are shown in table 5.

3.2.1 The FACT-P

In the worsened group, the score of the FACT-G and FACT-P total scales, FWB, PCS, and TOI decreased statistically significantly (deterioration in HRQOL), with all responsiveness statistics >0.2. In the improved group, all total and subscale scores, except for the SWB score, increased statistically significantly (improvement in HRQOL), with all responsiveness statistics >0.2. In the unchanged group, all total and subscale scores, except for the SWB and

FWB scores also increased statistically significantly but all responsiveness statistics in the unchanged group were smaller than those in the improved group.

The sensitivity analysis showed that all scores, except for the SWB score, increased statistically significantly in both improved group and unchanged group but all of the coefficients in the unchanged group, except for the PWB subscale, were smaller than those in the improved group. In contrast, all scales of the FACT-P did not detect any statistically significant difference in the worsened group.

3.2.2 The SF-12 v2

In the worsened group, the score of the PF, GH and SF domain and the PCS-12 decreased statistically significantly (deterioration in HRQOL), with all responsiveness statistics >0.2. In the improved group, all domain and summary scores of the SF-12 v2 increased statistically significantly improvement in HRQOL), with all responsiveness statistics >0.2. In the unchanged group, the score of the BP, VT, MH domains and MCS-12 also increased statistically significantly but all responsiveness statistics in the unchanged group were smaller than those in the improved group.

The sensitivity analysis showed that all SF-12 v2 domain scores and summary scores increased statistically significantly in the improved group. In the unchanged group, all summary and domain scores, except for the GH and SF domain scores also increased statistically significantly but all of the coefficients in the unchanged group, except for the RE domain, were smaller than those in the improved group. In contrast, all domains, PCS-12 and MCS-12 did not detect any statistically significant difference in the worsened group.

3.3 External responsiveness

Table 6 shows the difference in mean change and the AUC of ROC (i) between improved and unchanged group, (ii) between improved and worsened group, and (iii) between worsened and unchanged group.

Only the RP and PCS-12 of the SF-12 v2 could detect the difference in mean change between improved and unchanged group with the AUC standard of at least 0.70. The AUC of ROC was just short of 0.7 in the PWB (0.68), the TOI (0.69) and the FACT-P total scale (0.69). The FACT-P and SF-12 v2 (except the SWB of the FACT-P, and the BP, RE and MCS-12 of the SF-12 v2) could detect the difference in mean change between improved and worsened

group with the AUC standard of at least 0.70. The FACT-P (except the SWB and EWB) and only the PF of the SF-12 v2 could detect the difference in mean change between worsened and unchanged group with the AUC standard of at least 0.70.

Figure 2 shows the AUC of ROC of the PCS-12 and MCS-12 of the SF-12 v2, FACT-G and FACT-P total scales.

4. Discussion

To the best of our knowledge, this was the first study to focus on the assessment of the internal and external responsiveness of the SF-12 v2 and the FACT-P, which were recommended to evaluate the HRQOL of prostate cancer patients. It should be emphasized that it was not the aim of this study to compare the responsiveness of these two HRQOL measures. These measures are designed for completely different purposes. Furthermore, although the names of domains are similar (physical well-being in the FACT-P vs. physical functioning in the SF-12 v2), their respective contents are different substantially. A strong evidence for this was provided in table 2 where the PWB showed an increase but the PF a decrease over time. Therefore, a direct comparison is meaningless. Instead, these measures should be used complementarily.

Assessment of internal responsiveness by paired t-test and GEE confirmed that all scales of the FACT- P (except SWB subscale) and all domains and component summaries of the SF-12 v2 were responsive to positive change in subjects with improved general health. The SES and SRM in the improved group were the highest in the PWB subscale of the FACT-P, indicating that it was the most internally responsive to capture positive HRQOL changes in patients with prostate cancer, compared with SWB, EWB and FWB. However, the improvement detected by the PWB subscale should be interpreted with caution because in the sensitivity analysis, it was found that the unchanged group had a larger improvement in the PWB score than the improved group.

The sensitivity analysis suggested that the ability of the FACT-P and the SF-12 v2 to detect negative change in subjects with worsened general health was disappointing. The FACT-P and the SF-12 v2 were not responsive to negative change, overall. Our findings were contrary to the findings of another study which found that the PWB subscale was the most responsive to capture negative HRQOL change in Chinese patients with colorectal cancer [25]. There were some possible explanations. First, patients with worsened general health did not

necessarily have deteriorations in cancer-specific HRQOL. Second, compared with other cancer types, prostate cancer is relatively "indolent" with a slow progressive course and consequently, prostate cancer may have a less aggressive impact on patients' generic and cancer-specific HRQOL.

Noteworthy, the change detected by the FACT-P and SF-12 v2 should be interpreted with caution because positive changes were also detected in stable groups in some subscales/ domains even though the responsiveness statistics / coefficients in the stable group were smaller than those in the improved group. The measures might be too responsive. As a consequence, the positive changed detected in the stable group might be due to "noises", which are not clinically meaningful [25].

Compared with other subscales of the FACT-P, the internal responsiveness of the SWB was disappointing because it did not detect any change in neither improved group nor worsened group. One the contrary, a previous study in Chinese patients with colorectal cancer found that the SWB scale could detect positive change in "improved", "unchanged" and "worsened" group [25]. There were some possible explanations. First, the poorer internal responsiveness of the SWB subscale might be related to poor reproducibility [43] and the weak convergent validity of the SWB scale [23, 44]. Second, the question items such as "I feel close to my friends" and "I get support from my friends" might not be necessarily relevant in Chinese culture in which friends are not the primary support and recourse. Further studies are needed to evaluate and compare the content validity and relevance of the core module of the FACT in different cancer populations.

Concerning the external responsiveness, the FACT-P and the SF-12 v2 performed the best in distinguishing between subjects who had improved general health and those with worsened general health. Compared with the SF-12 v2, the FACT-P performed better in distinguishing between subjects who had unchanged general health and those with worsened general health. On the contrary, the SF-12 v2 performed slightly better in distinguishing between subjects who had unchanged general health and those with improved general health. Nevertheless, some of the AUC in the FACT-P were marginally acceptable, with the AUC just short of 0.7. Similar to the findings of internal responsiveness, the SWB had poor external responsiveness, with the AUC< 0.6 in three comparisons. The external responsiveness of the SWB was also problematic in Chinese patients with colorectal cancer [25].

5. Limitation

There were some limitations in the present study. First, the GRS might also be too generic. Thus, people with improved health status as measured by the GRS did not necessarily have improvement in disease-specific HRQOL, and vice versa [33]. However, the GRS was still the most commonly used external anchor to evaluate the responsiveness of instruments because of its validity and reliability. Besides, further study might use other cancer-specific HRQOL instruments as an external anchor. However, those questionnaires were lengthier, which might increase the burden of respondents. Second, the baseline data were collected by face-to-face interviews whilst the follow-up data were collected by telephone interviews. The difference in the administration mode between two interviews might lead to differences in scores. However, it was not feasible to request our study subjects to go to the hospital to finish the 6-month follow-up interview. In order to minimize the potential effects caused by the change in administration mode, the same interviewer conducted all the baseline and follow-up interviews. Furthermore, the interviewer was required to read the questionnaire verbatim in a standardized approach. Third, it was a convenience sample which might threaten the external validity of our study findings. Further study might use consecutive sampling methods to recruit study subjects. Finally, patients who were too ill to consent were excluded, which might lead to the low sample numbers in the "worsened group".

6. Conclusion

Assessment of internal responsiveness by paired t-test and GEE confirmed that all scales of the FACT- P (except SWB subscale) and all domains and component summaries of the SF-12 v2 were responsive to positive change in subjects with improved general health. On the contrary, the results of the GEE suggested that the FACT-P and the SF-12 v2 were not responsive to negative change. Besides, positive change detected by these measures should be interpreted with caution as they might be too responsive to detect "noise", which is not clinically significant. Concerning the external responsiveness, the FACT-P and the SF-12 v2 performed the best in distinguishing between improved l health and worsened health. The FACT-P performed better in distinguishing between unchanged health and worsened health. The SF-12 v2 performed better in distinguishing between unchanged health and improved health. These findings suggested that both generic and disease–specific measures should be used together to complement each other.

Word count: 3999

Abbreviations: HRQOL, Health-related Quality of Life; FACT-P, Functional Assessment of Cancer Therapy-Prostate; PWB, Physical well-being; SWB, Social well-being; EWB, Emotional well-being; FWB, Functional well-being; TOI, Trial Outcome Index; PF, Physical functioning; RP, Role physical; BP, Bodily pain; GH, General health; VT, Vitality; SF, Social functioning; RE, Role emotional; MH, Mental health; PCS-12, Physical composite summary; MCS-12, Mental composite summary; SES, Standardized effect size; SRM, Standardized response mean ; RS, Responsiveness statistic; GRS, Global Rating on Change Scale; SF-12 v2, Short Form-12 Health Survey version 2, ROC, Receiver operating characteristic; AUC, Area under the receiver operating characteristic curve; PSA, prostate specific antigen; GEE, Generalized estimating equation;

Funding: no funding

Conflict of interest: all authors declares that he/she has no conflict of interest

Ethics approval: The study protocol was approved by the institutional review boards: HKWC (Ref No.: UW13-239).All procedures performed in studies involving human participants were 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.

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

Reference

1. Ferlay, J., et al., Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. International Journal of Cancer, 2010. 127(12), 2893-2917.
2. Pu, Y.S., et al., Changing trends of prostate cancer in Asia. Aging Male, 2004. 7(2), 120-32.
3. Siegel, R., et al., Cancer statistics, 2014. CA: A Cancer Journal for Clinicians, 2014. 64(1), 9-29.
4. Sanda, M.G., et al., Quality of life and satisfaction with outcome among prostate-cancer survivors. N Engl J Med, 2008. 358(12), 1250-61.

5. Miller, D.C., et al., Long-term outcomes among localized prostate cancer survivors: health-related quality-of-life changes after radical prostatectomy, external radiation, and brachytherapy. J Clin Oncol, 2005. 23(12), 2772-80.
6. Litwin, M.S., et al., Quality-of-life outcomes in men treated for localized prostate cancer. JAMA, 1995. 273(2), 129-35.
7. Smith, D.P., et al., Quality of life three years after diagnosis of localised prostate cancer: population based cohort study. Bmj, 2009. 339.
8. Choi, E.P., et al., Health-related quality of life of Chinese patients with prostate cancer in comparison to general population and other cancer populations. Support Care Cancer, 2015.
9. Sommers, S.D. and S.D. Ramsey, A review of quality-of-life evaluations in prostate cancer. Pharmacoeconomics, 1999. 16(2), 127-40.
10. Hamoen, E.H., et al., Measuring health-related quality of life in men with prostate cancer: A systematic review of the most used questionnaires and their validity. Urol Oncol, 2015. 33(2), 69 e19-28.
11. Rnic, K., et al., Measuring symptoms in localized prostate cancer: a systematic review of assessment instruments. Prostate Cancer Prostatic Dis, 2013. 16(2), 111-22.
12. Revicki, D.A., et al., Recommendations on health-related quality of life research to support labeling and promotional claims in the United States. Qual Life Res, 2000. 9(8), 887-900.
13. Revicki, D.A., et al., Responsiveness and minimal important differences for patient reported outcomes. Health Qual Life Outcomes, 2006. 4, 70.
14. Testa, M.A. and J.F. Nackley, Methods for quality-of-life studies. Annu Rev Public Health, 1994. 15, 535-59.
15. Revicki, D., et al., Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. Journal of Clinical Epidemiology, 2008. 61(2), 102-9.
16. Guyatt, G., S. Walter, and G. Norman, Measuring change over time: assessing the usefulness of evaluative instruments. J Chronic Dis, 1987. 40(2), 171-8.
17. Husted, J.A., et al., Methods for assessing responsiveness: a critical review and recommendations. J Clin Epidemiol, 2000. 53(5), 459-68.

18. Choi, E.P., C.L. Lam, and W.-Y. Chin, The Incontinence Impact Questionnaire-7 (IIQ-7) can be applicable to chinese males and females with lower urinary tract symptoms. The Patient-Patient-Centered Outcomes Research, 2014. 7(4), 403-411.
19. Brazier, J.E. and J. Roberts, The Estimation of a Preference-Based Measure of Health From the SF-12. Medical Care, 2004. 42(9), 851-859.
20. Ware Jr, J.E., M. Kosinski, and S.D. Keller, A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Medical care, 1996. 34(3), 220-233.
21. 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? Journal of evaluation in clinical practice, 2013. 19(1), 200-208.
22. Esper, P., et al., Measuring quality of life in men with prostate cancer using the functional assessment of cancer therapy-prostate instrument. Urology, 1997. 50(6), 920-8.
23. Wong, C.K., et al., Psychometric properties of Functional Assessment of Cancer Therapy-Prostate (FACT-P) in Chinese patients with prostate cancer. Qual Life Res, 2015. 24(10), 2397-402.
24. Kamper, S.J., C.G. Maher, and G. Mackay, Global rating of change scales: a review of strengths and weaknesses and considerations for design. Journal of Manual & Manipulative Therapy, 2009. 17(3), 163-170.
25. Wong, C.K., et al., Condition-specific measure was more responsive than generic measure in colorectal cancer: all but social domains. J Clin Epidemiol, 2013. 66(5), 557-65.
26. Beaton, D.E., S. Hogg-Johnson, and C. Bombardier, Evaluating changes in health status: reliability and responsiveness of five generic health status measures in workers with musculoskeletal disorders. Journal of Clinical Epidemiology, 1997. 50(1), 79-93.
27. Wong, C.K., et al., Responsiveness was similar between direct and mapped SF-6D in colorectal cancer patients who declined. J Clin Epidemiol, 2014. 67(2), 219-27.
28. Juniper, E.F., et al., Determining a minimal important change in a disease-specific Quality of Life Questionnaire. J Clin Epidemiol, 1994. 47(1), 81-7.
29. Wright, J.G. and N.L. Young, A comparison of different indices of responsiveness. J Clin Epidemiol, 1997. 50(3), 239-46.

30. Stavem, K., S.S. Frøland, and K.B. Hellum, Comparison of preference-based utilities of the 15D, EQ-5D and SF-6D in patients with HIV/AIDS. Quality of Life Research, 2005. 14(4), 971-980.
31. Marra, C.A., et al., Are indirect utility measures reliable and responsive in rheumatoid arthritis patients? Quality of Life Research, 2005. 14(5), 1333-1344.
32. Gerhards, S.A., et al., The responsiveness of quality of life utilities to change in depression: a comparison of instruments (SF-6D, EQ-5D, and DFD). Value in Health, 2011. 14(5), 732-9.
33. 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.
34. Liang, M.H., A.H. Fossel, and M.G. Larson, Comparisons of Five Health Status Instruments for Orthopedic Evaluation. Medical Care, 1990. 28(7), 632-642.
35. Terwee, C.B., et al., On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. Quality of Life Research, 2003. 12(4), 349-362.
36. Guyatt, G., S. Walter, and G. Norman, Measuring change over time: assessing the usefulness of evaluative instruments. Journal of Chronic Diseases, 1987. 40(2), 171-8.
37. Cohen, J., *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. 1988, Hillsdale, NJ: Lawrence Erlbaum Associates.
38. Norman, G.R., P. Stratford, and G. Regehr, Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. Journal of Clinical Epidemiology, 1997. 50(8), 869-879.
39. Efron, B., Better Bootstrap Confidence Intervals. Journal of the American Statistical Association, 1987. 82(397), 171-185.
40. Rubin, D.B., *Multiple imputation for nonresponse in surveys*. Vol. 81. 2004: John Wiley & Sons.
41. Deyo, R.A. and R.M. Centor, Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. Journal of chronic diseases, 1986. 39(11), 897-906.
42. Terwee, C.B., et al., Quality criteria were proposed for measurement properties of health status questionnaires. Journal of Clinical Epidemiology, 2007. 60(1), 34-42.

43. Rotonda, C., et al., Validation of the French version of the colorectal-specific quality-of-life questionnaires EORTC QLQ-CR38 and FACT-C. Qual Life Res, 2008. 17(3), 437-45.
44. Wong, C.K., et al., Validity and reliability study on traditional Chinese FACT-C in Chinese patients with colorectal neoplasm. J Eval Clin Pract, 2012. 18(6), 1186-95.

Table 1. Baseline Characteristics of Prostate Cancer Patients

	Baseline (N=291)		Follow-up (N=168)		Attrition (N=123)	
	N	%	N	%	N	%
Demographic Characteristics						
Age, Mean (SD)**	74.9 (8.6)		72.9(8.0)		77.7 (8.7)	
Education ##						
No formal schooling	46	15.8%	16	9.5%	30	24.4%
Primary	98	33.7%	58	34.5%	40	32.5%
Secondary	90	30.9%	64	38.1%	26	21.1%
Tertiary or above	53	18.2%	28	16.7%	25	20.3%
Unknown	4	1.4%	2	1.2%	2	1.6%
Marital Status#						
Married	222	76.3%	138	82.1%	84	68.3%
Single	17	5.8%	10	6.0%	7	5.7%
Separated, divorced or widower	48	16.5%	18	10.7%	30	24.4%
Unknown	4	1.4%	2	1.2%	2	1.6%
Currently Working						
Yes	25	8.6%	20	11.9%	5	4.1%
No	262	90.0%	146	86.9%	116	94.3%
Unknown	4	1.4%	2	1.2%	2	1.6%
Monthly income (HKD$)						
≤20,000	238	81.8%	133	79.2%	105	85.4%
>20,000	49	16.8%	33	19.6%	16	13.0%
Unknown	4	1.4%	2	1.2%	2	1.6%
Clinical Characteristics						
PSA						
<0.1ng/ml	109	37.5%	70	41.7%	39	31.7%
≥0.1 & <10ng/ml	121	41.6%	60	35.7%	61	49.6%
≥10ng/ml	39	13.4%	24	14.3%	15	12.2%
Unknown	22	7.6%	14	8.3%	8	6.5%
AJCC Cancer Stage						
I	58	19.9%	27	16.1%	31	25.2%
II	75	25.8%	42	25.0%	33	26.8%
III	31	10.7%	20	11.9%	11	8.9%
IV	112	38.5%	70	41.7%	42	34.1%
Unknown	15	5.2%	9	5.4%	6	4.9%
Distant metastasis	59	20.3%	37	22.0%	22	17.9%
KPS ##						
Mean (SD)**	91.4 (12.5)		94.1 (8.7)		87.6 (15.6)	
≤70	23	7.9%	5	2.98%	18	14.6%
80	34	11.7%	16	9.52%	18	14.6%
90	64	22.0%	39	23.21%	25	20.3%
100	138	47.4%	90	53.57%	48	39.0%
Unknown	32	11.0%	18	10.71%	14	11.4%
Treatments						

Watchful waiting/ active surveillance	24	8.2%	12	7.1%	12	9.8%
Androgen deprivation/ Combined androgen blockade	117	40.2%	65	38.7%	52	42.3%
Radical prostatectomy#	102	35.1%	68	40.5%	34	27.6%
Radical curative radiation	45	15.5%	26	15.5%	19	15.4%
Adjuvant radiation	8	2.7%	4	2.4%	4	3.3%
Chemotherapy	4	1.4%	2	1.2%	2	1.6%
HRQOL scores	Mean (SD)		Mean (SD)		Mean (SD)	
Condition-specific, FACT-P						
PWB**	24.6 (3.6)		25.7 (2.6)		23.2 (4.3)	
SWB**	19.6 (5.2)		20.6 (4.3)		18.3 (6.0)	
EWB	21.1 (3.8)		21.4 (3.3)		20.7 (4.3)	
FWB**	19.7 (5.4)		20.9 (4.7)		18.1 (5.8)	
PCS**	35.9 ± (6.5)		37.5 (5.6)		33.7 (6.9)	
TOI**	80.2 ± (13.7)		84.0 (11.3)		75.0 (15.0)	
FACT-G Total score**	85.0 ± (13.9)		88.5 (11.6)		80.3 (15.4)	
FACT-P Total score**	120.9 ± (19.2)		126.0 (16.0)		114.0 (21.1)	
Generic, SF-12v2						
PF**	67.2 (34.7)		78.7 (28.4)		51.2 (36.3)	
RP**	71.1 (26.9)		78.9 (22.8)		60.5 (28.4)	
BP**	75.1 (27.6)		82.6 (22.2)		64.7 (30.9)	
GH**	59.4 (27.2)		67.1 (24.1)		48.8 (27.9)	
VT**	70.1 (24.6)		75.1 (23.3)		63.1 (24.7)	
SF**	76.7 (26.2)		82.9 (22.6)		68.2 (28.4)	
RE*	84.8 (22.2)		87.6 (19.4)		80.9 (25.2)	
MH**	81.3 (19.3)		84.4 (17.3)		76.8 (21.0)	
PCS-12**	43.4 (13.5)		47.8 (10.7)		37.3 (14.6)	
MCS-12	57.4 (9.9)		58.0 (8.9)		56.5 (11.2)	

Note:

SD=standard deviation; PSA=Prostate-specific antigen; KPS=Karnofsky performance status; AJCC=American Joint Committee on Cancer ; HRQOL=health-related quality of life; FACT-P subscales: PWB=physical well-being; SWB=social well-being; EWB=emotional well-being; FWB=functional well-being; PCS=prostate cancer subscale; TOI=trial outcome index; SF-12v2 subscales: PF=physical functioning; RP=role physical; BP=bodily pain; GH=general health; VT=vitality; SF=social functioning; RE=role emotional; MH=mental health; PCS-12=physical component summary 12; MCS-12=mental component summary 12;

* : p-value <0.05 by independent t-test

** : p-value <0.01 by independent t-test

: p-value <0.05 by chi-square test

: p-value <0.01 by chi-square test

Table 2. Baseline, 6-month Follow-up and Mean Change on the Condition-specific and Generic HRQOL scores of Patients (n=168)

Measure/Subscale†	Baseline (SD)	6-month (SD)	Mean Change (SD)	P-value
Condition-specific, FACT-P				
PWB (range 0-28)	25.68 (2.61)	26.66 (2.60)	0.98 (2.26)	< 0.001*
SWB (range 0-24)	20.56 (4.29)	20.75 (4.71)	0.19 (3.41)	0.464
EWB (range 0-28)	21.38 (3.31)	22.40 (2.89)	1.03 (2.30)	< 0.001*
FWB (range 0-28)	20.87 (4.71)	21.20 (5.03)	0.33 (3.56)	0.235
PCS (range 0-48)	37.49 (5.63)	39.17 (5.04)	1.68 (4.52)	< 0.001*
TOI (range 0-104)	84.04 (11.33)	87.02 (11.59)	2.99 (8.60)	< 0.001*
FACT-G Total score (range 0-108)	88.49 (11.57)	91.02 (12.34)	2.53 (7.41)	< 0.001*
FACT-P Total score (range 0-156)	125.98 (16.01)	130.18 (16.70)	4.21 (10.84)	< 0.001*
Generic, SF-12v2				
PF (range 0-100)	78.72 (28.38)	77.83 (29.65)	-0.89 (25.06)	0.645
RP (range 0-100)	78.87 (22.85)	84.00 (21.41)	5.13 (22.37)	0.003*
BP (range 0-100)	82.59 (22.19)	90.48 (17.90)	7.89 (20.50)	< 0.001*
GH (range 0-100)	67.11 (24.06)	65.15 (23.57)	-1.96 (20.03)	0.205
VT (range 0-100)	75.15 (23.29)	88.39 (19.32)	13.24 (21.87)	< 0.001*
SF (range 0-100)	82.89 (22.58)	82.89 (19.56)	0.00 (19.54)	1.000
RE (range 0-100)	87.65 (19.39)	91.00 (18.20)	3.35 (17.06)	0.012*
MH (range 0-100)	84.45 (17.28)	90.25 (16.33)	5.80 (16.69)	< 0.001*
PCS-12 mean :50(SD:10) #	47.83 (10.67)	48.44 (9.86)	0.61 (8.58)	0.355
MCS-12 mean :50(SD:10)#	57.99 (8.85)	60.91 (8.54)	2.92 (7.43)	< 0.001*

Note:
HRQOL=health-related quality of life; SD=standard deviation; FACT-P subscales: PWB=physical well-being; SWB=social well-being; EWB=emotional well-being; FWB=functional well-being; PCS=prostate cancer subscale; TOI=trial outcome index; SF-12v2 subscales: PF=physical functioning; RP=role physical; BP=bodily pain; GH=general health; VT=vitality; SF=social functioning; RE=role emotional; MH=mental health; PCS-12=physical component summary 12; MCS-12=mental component summary 12

* Significant Difference on HRQOL between Baseline and 6-month follow-up

† Higher scores represents a higher level of functioning or a better HRQOL.

mean 50 with SD :10 due to norm-based scoring

Table 3. Distribution of Global Rating on Change Scale

Response		Follow-up (n=168)
-2	worse	3 (1.79%)
-1	a little worse	11 (6.55%)
0	same	120 (71.43%)
1	a little better	31 (18.45%)
2	better	3 (1.79%)

Table 4. Mean Change, Standardized Effect Size, Standardized Response Mean and Responsiveness Statistic of HRQOL Scores by Global Rating on Change Scale (n=168)

Measure/Subscale†	Mean (SD) at baseline	Mean (SD) at 6-month follow-up	Mean Change (SD)	P-value	SES (95% CI)	SRM (95% CI)	RS (95% CI)
Worsened group (n=14)							
Condition-specific, FACT-P							
PWB	24.00 (3.21)	22.21 (5.48)	-1.79 (4.63)	0.172	-0.33 (-0.70,0.10)	-0.39 (-0.79,0.19)	-1.09 (-3.23,0.08)
SWB	19.77 (5.56)	19.00 (5.14)	-0.77 (4.06)	0.489	-0.15 (-0.80,0.27)	-0.19 (-0.70,0.46)	-0.22 (-0.91,0.33)
EWB	19.71 (5.08)	19.07 (5.95)	-0.64 (3.97)	0.555	-0.11 (-0.56,0.12)	-0.16 (-0.67,0.56)	-0.34 (-2.62,0.34)
FWB	19.21 (4.37)	14.86 (5.83)	-4.36 (4.07)	0.001*	-0.75 (-1.18,-0.40)	-1.07 (-1.68,-0.55)	-1.34 (-2.11,-0.69)
PCS	35.57 (4.50)	31.64 (7.97)	-3.93 (6.59)	0.044*	-0.49 (-0.77,-0.20)	-0.60 (-0.89,-0.24)	-1.30 (-3.16,-0.43)
TOI	78.79 (10.07)	68.71 (18.32)	-10.07 (13.66)	0.016*	-0.55 (-0.80,-0.33)	-0.74 (-0.99,-0.45)	-1.70 (-3.66,-0.79)
FACT-G Total score	82.70 (13.35)	75.14 (18.00)	-7.56 (11.07)	0.024*	-0.42 (-0.73,-0.10)	-0.68 (-1.05,-0.03)	-1.27 (-2.49,-0.43)
FACT-P Total score	118.27 (16.46)	106.79 (24.92)	-11.49 (16.79)	0.024*	-0.46 (-0.77,-0.17)	-0.68 (-0.95,-0.15)	-1.50 (-3.15,-0.56)
Generic, SF-12v2							
PF	67.86 (37.25)	42.86 (33.15)	-25.00 (40.43)	0.038*	-0.75 (-1.59,0.08)	-0.62 (-1.37,0.19)	-1.22 (-2.20,0.00)
RP	67.86 (27.17)	58.93 (18.62)	-8.93 (25.68)	0.216	-0.48 (-1.14,0.18)	-0.35 (-0.84,0.24)	-0.44 (-1.35,0.14)
BP	75.00 (27.74)	73.21 (24.93)	-1.79 (31.72)	0.836	-0.07 (-0.68,0.71)	-0.06 (-0.60,0.57)	-0.10 (-1.20,0.74)
GH	49.64 (27.14)	29.64 (26.99)	-20.00 (29.61)	0.025*	-0.74 (-1.57,-0.12)	-0.68 (-1.15,0.02)	-1.17 (-2.27,-0.38)
VT	69.64 (22.31)	71.43 (27.49)	1.79 (28.53)	0.818	0.06 (-0.65,0.55)	0.06 (-0.56,0.66)	0.09 (-0.93,0.78)
SF	71.43 (27.49)	57.14 (20.64)	-14.29 (18.90)	0.014*	-0.69 (-1.13,-0.12)	-0.76 (-1.36,0.00)	-0.86 (-1.48,-0.21)
RE	81.25 (25.36)	73.21 (28.53)	-8.04 (25.76)	0.264	-0.28 (-0.78,0.11)	-0.31 (-0.72,0.27)	-0.50 (-1.81,0.18)
MH	81.25 (20.07)	76.79 (21.29)	-4.46 (21.15)	0.444	-0.21 (-0.78,0.32)	-0.21 (-0.74,0.39)	-0.31 (-1.16,0.48)
PCS-12	42.49 (14.30)	35.70 (9.18)	-6.79 (11.72)	0.049*	-0.74 (-1.45,-0.02)	-0.58 (-1.08,-0.02)	-1.00 (-2.10,-0.27)
MCS-12	56.16 (11.25)	55.16 (14.03)	-1.01 (11.00)	0.737	-0.07 (-0.52,0.34)	-0.09 (-0.61,0.53)	-0.17 (-1.35,0.67)
Unchanged group (n=120)							
Condition-specific, FACT-P							
PWB	25.94 (2.54)	26.94 (1.83)	1.00 (1.64)	< 0.001*	0.55 (0.36,0.75)	0.61 (0.40,0.77)	0.61 (0.40,0.77)
SWB	20.88 (4.25)	21.08 (4.79)	0.20 (3.48)	0.539	0.04 (-0.09,0.16)	0.06 (-0.12,0.23)	0.06 (-0.13,0.24)
EWB	21.78 (2.91)	22.75 (2.22)	0.97 (1.91)	< 0.001*	0.44 (0.25,0.66)	0.51 (0.36,0.63)	0.51 (0.35,0.64)

FWB	21.28 (5.00)	21.74 (4.81)	0.46 (3.25)	0.125	0.10 (-0.03,0.23)	0.14 (-0.05,0.30)	0.14 (-0.04,0.31)
PCS	38.24 (5.27)	39.87 (4.14)	1.63 (3.03)	< 0.001*	0.39 (0.26,0.55)	0.54 (0.35,0.72)	0.54 (0.34,0.71)
TOI	85.47 (11.27)	88.55 (9.70)	3.08 (5.93)	< 0.001*	0.32 (0.20,0.44)	0.52 (0.32,0.70)	0.52 (0.32,0.70)
FACT-G Total score	89.89 (11.53)	92.51 (11.31)	2.62 (5.96)	< 0.001*	0.23 (0.14,0.34)	0.44 (0.27,0.62)	0.44 (0.25,0.62)
FACT-P Total score	128.13 (15.69)	132.38 (14.75)	4.25 (7.64)	< 0.001*	0.29 (0.19,0.39)	0.56 (0.37,0.74)	0.56 (0.35,0.75)
Generic, SF-12v2							
PF	81.67 (26.48)	79.38 (28.38)	-2.29 (20.50)	0.223	-0.08 (-0.20,0.05)	-0.11 (-0.28,0.07)	-0.11 (-0.27,0.08)
RP	81.88 (21.67)	84.06 (21.68)	2.19 (20.22)	0.238	0.10 (-0.08,0.28)	0.11 (-0.08,0.31)	0.11 (-0.09,0.31)
BP	84.38 (21.27)	91.67 (17.55)	7.29 (18.73)	< 0.001*	0.42 (0.20,0.70)	0.39 (0.23,0.52)	0.39 (0.24,0.53)
GH	69.79 (22.71)	66.88 (21.88)	-2.92 (17.12)	0.064	-0.13 (-0.27,0.01)	-0.17 (-0.35,0.02)	-0.17 (-0.35,0.02)
VT	77.92 (23.19)	89.79 (18.19)	11.88 (18.89)	< 0.001*	0.65 (0.44,0.94)	0.63 (0.48,0.79)	0.63 (0.47,0.77)
SF	86.25 (20.71)	84.38 (18.64)	-1.88 (16.58)	0.218	-0.10 (-0.25,0.05)	-0.11 (-0.27,0.06)	-0.11 (-0.28,0.08)
RE	88.13 (18.96)	91.98 (17.13)	3.85 (16.03)	0.010*	0.22 (-0.01,0.44)	0.24 (-0.03,0.44)	0.24 (-0.02,0.43)
MH	86.35 (16.20)	91.15 (15.99)	4.79 (14.33)	< 0.001*	0.30 (0.12,0.53)	0.33 (0.14,0.51)	0.33 (0.16,0.51)
PCS-12	49.06 (9.77)	48.86 (9.73)	-0.20 (6.80)	0.750	-0.02 (-0.14,0.10)	-0.03 (-0.22,0.14)	-0.03 (-0.19,0.16)
MCS-12	58.66 (8.60)	61.42 (7.86)	2.75 (6.07)	< 0.001*	0.35 (0.19,0.52)	0.45 (0.26,0.62)	0.45 (0.26,0.62)
Improved group (n=34)							
Condition-specific, FACT-P							
PWB	25.44 (2.35)	27.50 (0.90)	2.06 (1.82)	< 0.001*	2.30 (1.65,3.37)	1.13 (0.83,1.43)	1.26 (0.87,1.75)
SWB	19.75 (3.83)	20.33 (4.12)	0.58 (2.86)	0.244	0.14 (-0.10,0.35)	0.20 (-0.18,0.52)	0.17 (-0.13,0.44)
EWB	20.65 (3.52)	22.56 (2.31)	1.91 (2.31)	< 0.001*	0.83 (0.25,1.63)	0.83 (0.61,1.10)	1.00 (0.57,1.50)
FWB	20.09 (3.47)	21.88 (3.52)	1.79 (2.80)	< 0.001*	0.51 (0.27,0.82)	0.64 (0.33,0.91)	0.55 (0.27,0.90)
PCS	35.62 (6.73)	39.80 (3.95)	4.18 (5.79)	< 0.001*	1.06 (0.48,2.06)	0.72 (0.43,1.04)	1.38 (0.92,2.41)
TOI	81.15 (11.10)	89.18 (7.27)	8.03 (8.70)	< 0.001*	1.11 (0.68,1.78)	0.92 (0.57,1.41)	1.36 (0.95,2.12)
FACT-G Total score	85.93 (9.95)	92.27 (8.05)	6.35 (6.58)	< 0.001*	0.79 (0.53,1.19)	0.97 (0.54,1.37)	1.06 (0.70,1.51)
FACT-P Total score	121.54 (15.52)	132.07 (11.33)	10.53 (11.28)	< 0.001*	0.93 (0.59,1.49)	0.93 (0.55,1.43)	1.38 (0.96,2.08)
Generic, SF-12v2							
PF	72.79 (29.75)	86.76 (22.39)	13.97 (23.19)	0.001*	0.62 (0.30,1.07)	0.60 (0.32,0.85)	0.68 (0.31,1.16)
RP	72.79 (23.12)	94.12 (10.76)	21.32 (20.30)	< 0.001*	1.98 (1.30,3.01)	1.05 (0.74,1.40)	1.05 (0.62,1.53)

BP	79.41 (22.59)	93.38 (11.20)	13.97 (19.65)	< 0.001*	1.25 (0.72,2.05)	0.71 (0.43,1.02)	0.75 (0.38,1.18)
GH	64.85 (24.88)	73.68 (13.33)	8.82 (19.15)	0.011*	0.66 (0.24,1.22)	0.46 (0.17,0.76)	0.52 (0.16,0.94)
VT	67.65 (22.64)	90.44 (16.30)	22.79 (25.65)	< 0.001*	1.40 (0.67,2.37)	0.89 (0.53,1.26)	1.21 (0.72,1.76)
SF	75.74 (24.22)	88.24 (14.08)	12.50 (23.23)	0.004*	0.89 (0.35,1.58)	0.54 (0.23,0.84)	0.75 (0.31,1.28)
RE	88.60 (18.30)	94.85 (11.96)	6.25 (14.84)	0.019*	0.52 (0.14,1.25)	0.42 (0.18,0.65)	0.39 (0.09,0.83)
MH	79.04 (18.90)	92.65 (12.73)	13.60 (19.55)	< 0.001*	1.07 (0.46,1.97)	0.70 (0.35,1.00)	0.95 (0.51,1.48)
PCS-12	45.66 (11.38)	52.19 (5.76)	6.53 (9.51)	< 0.001*	1.13 (0.58,1.80)	0.69 (0.42,0.96)	0.96 (0.48,1.53)
MCS-12	56.36 (8.64)	61.48 (7.30)	5.13 (9.33)	0.003*	0.70 (0.22,1.47)	0.55 (0.22,0.90)	0.84 (0.36,1.42)

Note:
HRQOL=health-related quality of life; SD=standard deviation; FACT-P subscales: PWB=physical well-being; SWB=social well-being; EWB=emotional well-being; FWB=functional well-being; PCS=prostate cancer subscale; TOI=trial outcome index; SF-12v2 subscales: PF=physical functioning; RP=role physical; BP=bodily pain; GH=general health; VT=vitality; SF=social functioning; RE=role emotional; MH=mental health; PCS-12=physical component summary 12; MCS-12=mental component summary 12; SES=standardized effect size; SRM=standardized response mean; RS=responsiveness statistic

* Significant Difference on HRQOL between Baseline and 6-month follow-up

† Higher scores represents a higher level of functioning or a better HRQOL.

Table 5. Sensitivity analysis of HRQOL Scores using generalized estimating equations by Global Rating on Change Scale after multiple imputation

Measure/Subscale†	Worsened group		Unchanged group		Improved group	
	Coefficient (95% CI)	P-value	Coefficient (95% CI)	P-value	Coefficient (95% CI)	P-value
Condition-specific, FACT-P						
PWB	0.57 (-2.58,3.73)	0.722	1.94 (1.38,2.50)	< 0.001*	1.74 (0.91,2.57)	< 0.001*
SWB	0.13 (-1.96,2.22)	0.904	0.83 (-0.06,1.71)	0.067	0.73 (-0.69,2.15)	0.312
EWB	1.33 (-1.71,4.37)	0.390	1.26 (0.72,1.80)	< 0.001*	1.54 (0.69,2.38)	< 0.001*
FWB	-2.24 (-5.32,0.85)	0.155	1.33 (0.50,2.17)	0.002*	1.49 (0.47,2.50)	0.004*
PCS	-1.21 (-5.43,3.00)	0.572	2.94 (2.00,3.88)	< 0.001*	3.33 (1.36,5.31)	0.001*
TOI	-2.88 (-12.73,6.97)	0.567	6.21 (4.21,8.22)	< 0.001*	6.57 (3.36,9.77)	< 0.001*
FACT-G Total score	-0.20 (-9.72,9.32)	0.966	5.36 (3.30,7.42)	< 0.001*	5.50 (2.72,8.29)	< 0.001*
FACT-P Total score	-1.42 (-14.83,11.99)	0.836	8.30 (5.54,11.06)	< 0.001*	8.84 (4.47,13.20)	< 0.001*
Generic, SF-12v2						
PF	-13.01 (-35.61,9.58)	0.259	8.30 (3.04,13.56)	0.002*	13.26 (4.13,22.39)	0.004*
RP	2.54 (-14.45,19.52)	0.770	9.64 (5.13,14.16)	< 0.001*	19.97 (12.84,27.10)	< 0.001*
BP	5.66 (-11.46,22.78)	0.517	13.70 (9.52,17.88)	< 0.001*	14.82 (7.77,21.86)	< 0.001*
GH	-7.78 (-26.47,10.91)	0.415	3.54 (-0.64,7.73)	0.097	7.94 (1.23,14.64)	0.020*
VT	6.99 (-7.83,21.80)	0.356	17.10 (13.30,20.90)	< 0.001*	23.48 (14.64,32.31)	< 0.001*

SF	-6.10 (-20.61,8.40)	0.410	3.92 (-0.12,7.96)	0.057	10.73 (3.22,18.24)	0.005*
RE	2.43 (-14.99,19.85)	0.785	6.23 (2.70,9.76)	0.001*	5.40 (0.15,10.64)	0.044*
MH	3.31 (-9.70,16.32)	0.618	8.82 (5.78,11.86)	< 0.001*	11.66 (5.67,17.65)	< 0.001*
PCS-12	-2.55 (-9.95,4.85)	0.500	3.69 (1.71,5.66)	< 0.001*	6.53 (3.10,9.96)	< 0.001*
MCS-12	2.32 (-4.76,9.41)	0.520	3.64 (2.07,5.22)	< 0.001*	4.43 (1.44,7.42)	0.004*

Note: HRQOL=health-related quality of life; SD=standard deviation; FACT-P subscales: PWB=physical well-being; SWB=social well-being; EWB=emotional well-being; FWB=functional well-being; PCS=prostate cancer subscale; TOI=trial outcome index; SF-12v2 subscales: PF=physical functioning; RP=role physical; BP=bodily pain; GH=general health; VT=vitality; SF=social functioning; RE=role emotional; MH=mental health; PCS-12=physical component summary 12; MCS-12=mental component summary 12; SES=standardized effect size; SRM=standardized response mean; RS=responsiveness statistic

* Significant Difference on HRQOL between Baseline and 6-month follow-up

† Higher scores represent a higher level of functioning or a better HRQOL.

Table 6. Mean Change and Area under the Receiver Operating Characteristic Curve on Discriminating Subjects with Worsened/Unchanged and Improved Health Status

Measure/Subscale†	Mean difference (95% CI)	AUC (95% CI)*	Correlation (95% CI)
Improved Vs Unchanged			
Condition-specific, FACT-P			
PWB	-1.06 (-1.70,-0.41)	0.68 (0.59,0.78)	0.25 (0.10,0.40)
SWB	-0.39 (-1.68,0.90)	0.52 (0.42,0.62)	0.05 (-0.11,0.20)
EWB	-0.94 (-1.71,-0.17)	0.62 (0.51,0.72)	0.19 (0.04,0.34)
FWB	-1.34 (-2.55,-0.12)	0.63 (0.54,0.73)	0.17 (0.02,0.32)
PCS	-2.55 (-4.01,-1.09)	0.65 (0.55,0.75)	0.27 (0.12,0.41)
TOI	-4.95 (-7.49,-2.41)	0.69 (0.60,0.79)	0.30 (0.15,0.44)
FACT-G Total score	-3.72 (-6.07,-1.38)	0.67 (0.57,0.77)	0.25 (0.09,0.39)
FACT-P Total score	-6.28 (-9.57,-2.99)	0.69 (0.59,0.79)	0.29 (0.14,0.43)
Generic, SF-12v2			
PF	-16.26 (-24.36,-8.16)	0.66 (0.58,0.74)	0.31 (0.16,0.44)
RP	-19.14 (-26.90,-11.37)	0.73 (0.63,0.82)	0.37 (0.22,0.50)
BP	-6.68 (-13.95,0.59)	0.60 (0.50,0.70)	0.15 (-0.01,0.30)
GH	-11.74 (-18.49,-4.99)	0.65 (0.55,0.74)	0.27 (0.12,0.41)
VT	-10.92 (-18.81,-3.03)	0.64 (0.53,0.74)	0.22 (0.06,0.36)
SF	-14.38 (-21.37,-7.38)	0.67 (0.57,0.76)	0.31 (0.16,0.45)
RE	-2.40 (-8.45,3.66)	0.51 (0.41,0.60)	0.06 (-0.10,0.22)
MH	-8.81 (-14.80,-2.82)	0.62 (0.51,0.73)	0.23 (0.07,0.37)
PCS-12	-6.72 (-9.59,-3.86)	0.70 (0.59,0.80)	0.35 (0.21,0.48)
MCS-12	-2.37 (-5.02,0.28)	0.57 (0.45,0.69)	0.14 (-0.02,0.29)
Improved Vs Worsened			
Condition-specific, FACT-P			
PWB	-3.84 (-5.70,-1.99)	0.82 (0.67,0.97)	0.52 (0.28,0.70)
SWB	-1.36 (-3.43,0.72)	0.59 (0.40,0.78)	0.19 (-0.10,0.45)
EWB	-2.55 (-4.40,-0.71)	0.70 (0.55,0.86)	0.38 (0.11,0.60)
FWB	-6.15 (-8.20,-4.10)	0.89 (0.78,0.99)	0.66 (0.47,0.80)
PCS	-8.11 (-11.96,-4.25)	0.93 (0.84,1.00)	0.53 (0.29,0.71)
TOI	-18.10 (-24.71,-11.49)	0.98 (0.96,1.00)	0.63 (0.42,0.78)
FACT-G Total score	-13.91 (-19.09,-8.73)	0.89 (0.78,1.00)	0.62 (0.41,0.77)
FACT-P Total score	-22.02 (-30.37,-13.66)	0.94 (0.86,1.00)	0.62 (0.40,0.77)
Generic, SF-12v2			
PF	-38.97 (-57.58,-20.36)	0.82 (0.67,0.98)	0.53 (0.29,0.71)
RP	-30.25 (-44.29,-16.22)	0.82 (0.69,0.94)	0.54 (0.30,0.71)
BP	-15.76 (-30.90,-0.61)	0.64 (0.47,0.81)	0.30 (0.01,0.53)
GH	-28.82 (-43.27,-14.38)	0.79 (0.64,0.94)	0.51 (0.26,0.69)
VT	-21.01 (-37.94,-4.07)	0.71 (0.55,0.86)	0.35 (0.07,0.57)
SF	-26.79 (-40.91,-12.66)	0.81 (0.68,0.93)	0.49 (0.24,0.68)
RE	-14.29 (-26.17,-2.40)	0.64 (0.48,0.80)	0.34 (0.06,0.57)
MH	-18.07 (-30.86,-5.27)	0.73 (0.57,0.89)	0.39 (0.12,0.60)
PCS-12	-13.31 (-19.82,-6.80)	0.80 (0.64,0.96)	0.52 (0.28,0.70)

MCS-12	-6.13 (-12.42,0.15)	0.64 (0.46,0.82)	0.28 (-0.01,0.52)
Worsened Vs Unchanged			
Condition-specific, FACT-P			
PWB	-2.79 (-3.98,-1.60)	0.70 (0.52,0.88)	0.37 (0.22,0.51)
SWB	-0.97 (-2.95,1.01)	0.56 (0.40,0.73)	0.08 (-0.09,0.25)
EWB	-1.61 (-2.84,-0.38)	0.60 (0.43,0.76)	0.22 (0.05,0.38)
FWB	-4.82 (-6.68,-2.95)	0.81 (0.68,0.95)	0.41 (0.25,0.54)
PCS	-5.55 (-7.53,-3.57)	0.83 (0.72,0.94)	0.43 (0.29,0.56)
TOI	-13.15 (-17.11,-9.20)	0.89 (0.82,0.96)	0.50 (0.36,0.61)
FACT-G Total score	-10.18 (-13.89,-6.47)	0.81 (0.67,0.95)	0.43 (0.28,0.56)
FACT-P Total score	-15.74 (-20.75,-10.73)	0.86 (0.75,0.97)	0.48 (0.33,0.60)
Generic, SF-12v2			
PF	-22.71 (-35.69,-9.73)	0.74 (0.57,0.90)	0.29 (0.12,0.44)
RP	-11.12 (-22.75,0.51)	0.63 (0.47,0.80)	0.16 (-0.01,0.32)
BP	-9.08 (-20.46,2.31)	0.55 (0.38,0.72)	0.14 (-0.03,0.30)
GH	-17.08 (-27.54,-6.62)	0.69 (0.52,0.85)	0.27 (0.11,0.42)
VT	-10.09 (-21.29,1.11)	0.59 (0.44,0.75)	0.15 (-0.02,0.31)
SF	-12.41 (-21.81,-3.01)	0.69 (0.54,0.84)	0.22 (0.05,0.38)
RE	-11.89 (-21.52,-2.26)	0.63 (0.48,0.79)	0.21 (0.04,0.36)
MH	-9.26 (-17.71,-0.80)	0.64 (0.46,0.82)	0.19 (0.02,0.34)
PCS-12	-6.59 (-10.74,-2.44)	0.67 (0.47,0.86)	0.26 (0.10,0.41)
MCS-12	-3.76 (-7.51,-0.01)	0.57 (0.37,0.76)	0.17 (0.00,0.33)

Note:
AUC=Area under the receiver operating characteristic curve; HRQOL=health-related quality of life; SD=standard deviation; FACT-P subscales: PWB=physical well-being; SWB=social well-being; EWB=emotional well-being; FWB=functional well-being; PCS=prostate cancer subscale; TOI=trial outcome index; SF-12v2 subscales: PF=physical functioning; RP=role physical; BP=bodily pain; GH=general health; VT=vitality; SF=social functioning; RE=role emotional; MH=mental health; PCS-12=physical component summary 12; MCS-12=mental component summary 12

* The AUC $\geq$ 0.7 was considered adequate.

† Higher scores represent a higher level of functioning or a better HRQOL.